

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Traffic Management in Heterogeneous Mobile Networks

Jiang, Menglan

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Traffic Management in Heterogeneous Mobile Networks

Author:

Menglan Jiang

Supervisor:

Dr Mahmoodi Toktam

*A Thesis Submitted for the Degree of
Doctor of Philosophy
at
Center for Telecommunications Research
King's College London*



London, December 8, 2016

Traffic Management in Heterogeneous Mobile Networks

by

Menglan Jiang

A Thesis Submitted for the Degree of

Doctor of Philosophy

at

King's College London

December 8, 2016

Acknowledgement

First and foremost, I would like to express my deepest gratitude to my supervisor Dr. Mahmoodi Toktam for her continuous support, insightful guidance, and inspirational mentorship. Thank you for everything you did for me and for always encouraging me during my PhD study at King's College London.

Special thanks to Massimo Condoluci and Luis Guijarro, thank you for your great cooperation, helpful guidance and encouraging me to continue my research. I would also like to thank all the people I met in the Center of Telecommunications Research (CTR), including Jing An, Bi Zhao, Yuhang Xu, Changtao Zhong, Jinwei Gang, Maria, Adnan, Christoforos and Giorgos. Because you all, the time I was staying at CTR will be the most memorable time of my life.

Special thanks to my best friend Letao Xu and Yue Liu, thank you for always supporting me and encouraging me to overcome all difficulties.

Last, but not the least, I would like to thanks my family, who has always stood by me and believe in me. It is because of them, I own all the success in my life.

Abstract

Enhancing the efficiency of traffic management and improving the network revenue for operators are the most essential issues in heterogeneous mobile networks. As the heterogeneous mobile network is composed of diverse network technologies, it is expected to offer effective network connections and sufficient network resources to increased number of mobile users with various traffic demands. Moreover, the price of each selling resources might vary significantly from the business point of view. Therefore, heterogeneous mobile networks aim to provide superior user experience, guarantee efficient resource utilisation and maximise mobile network revenue.

The main objective of this thesis is to manage and allocate limited network resources to mobile users for improving their quality level and network revenue. The primary focus of this thesis is on developing valid network architecture for efficient resource allocation and designing novel optimisation algorithms for effective traffic management. Besides, the other focus of this thesis is on exploring future works in the field of traffic management in heterogeneous mobile networks.

This thesis covers two aspects of traffic management from the mobile user side and the network side. With the motivation of ever increasing traffic demands in heterogeneous mobile networks, maintaining quality level of mobile users, improving resource utilisation and maximising network revenue are more than ever challenging. In this thesis, a fully distributed device-

controlled network assisted mechanism for network selection has been developed. Decisions in this mechanism are made by mobile devices only. The device controlled mechanism mainly focuses on Quality of Service (QoS) requirements of mobile devices, while network information and location profiles are provided by the network assisted part. In order to improve the network performance, a novel semi-centralised QoS-based traffic management mechanism has been developed further. The semi-centralised mechanism aims to maintain the quality level of mobile users and guarantee the fairness among mobile users in heterogeneous mobile networks.

Moreover, a novel prioritised admission control mechanism has been developed for network slicing. Network slicing, where virtualised network resources are packaged and assigned in an isolated manner to set of mobile users, is considered as a key paradigm to implement resource allocation based on diversity of resource requirements. The mechanism is heuristic based and contains two-tier priority level among different service slices and mobile users. The mechanism aims to maximise resource utilisation and guarantee the mobile users resource requirements. Apart from this, network slicing is also significant for the business model of operators. As a result, a novel economics based traffic management mechanism has been developed further. It aims to maximise the network revenue and enhance network utilisation in heterogeneous mobile networks.

Finally, the thesis is concluded and some proper future researches are included as well.

Contents

1	Introduction	13
1.1	Introduction	13
1.2	Research Contributions	14
1.3	Publications	16
1.4	Organisation of this Thesis	17
2	Background	19
2.1	Introduction	19
2.2	Traffic Management Architectures	22
2.2.1	Distributed Device/User-Centric Architecture	22
2.2.2	Network Assisted Archtiecture	25
2.3	Device Centric Network Assisted Traffic Management	29
2.3.1	Multi-Attribute Decision Making	30
2.3.2	Reinforcement Learning	32
2.4	Novel Traffic Management in the 5G Mobile Networks	35
2.4.1	The Future Network Technology: 5G	35
2.4.2	Network Virtualization in 5G	37
2.4.3	Network Slicing in the 5G Mobile Networks	40
2.4.4	Admission Control in Virtualised Mobile Networks	42
2.4.5	Auction Model in Virtualised Mobile Networks	44

CONTENTS

3 Distributed Traffic Management in the Heterogeneous Mobile Networks	47
3.1 Introduction	47
3.2 Related Work	49
3.3 System Model	51
3.3.1 Model of the System-level Architecture	51
3.3.2 UE's Battery Models	52
3.4 Device Controlled Network Selection	54
3.4.1 Multi-Attribute Decision Making Mechanism	55
3.4.1.1 TOPSIS	55
3.4.1.2 Network Analytics	58
3.4.1.3 Location Profile and Location Estimation	59
3.4.2 Utility Definition	59
3.4.3 Reinforcement Learning	61
3.5 Performance Investigations	62
3.5.1 MADM Mechanism	62
3.5.1.1 Simulation Cases	64
3.5.1.2 Results Analysis	67
3.5.2 Reinforcement Learning Mechanism	69
3.5.2.1 Simulation Cases	70
3.5.2.2 Results Analysis	70
3.6 Conclusions	72
4 Fairness Based Traffic Management in 5G	74
4.1 Introduction	74
4.2 Related Works	76
4.3 System Model	77
4.4 Problem Formulation	80
4.4.1 Selfish Users	80

CONTENTS

4.4.2	Fair Network	81
4.5	Traffic Management Solutions	82
4.5.1	Solutions to (P1) Using Q-learning	82
4.5.2	Solutions to (P2) Using Simulated Annealing	83
4.6	Performance Investigations	85
4.6.1	Simulation Parameters	85
4.6.2	Simulation Cases	87
4.6.3	Result Analysis	88
4.7	Conclusions	95
5	Network Slicing based Traffic Management in 5G	96
5.1	Introduction	96
5.2	Related Works	99
5.3	System Model	101
5.3.1	Service Slices	102
5.3.2	Virtual Network	102
5.3.3	Physical resources	103
5.4	Two-tier Admission Control and Resource Allocation	105
5.4.1	Admission Control Strategy	105
5.4.2	Resource Allocation	109
5.5	Performance Investigation	111
5.5.1	Simulation Cases	112
5.5.2	Result Analysis	113
5.6	Conclusions	117
6	An Auction based Network Slicing in 5G Mobile Networks	118
6.1	Introduction	118
6.2	Related Works	121
6.3	System Model	124

CONTENTS

6.3.1	Network Slices	124
6.3.2	Network Resources	125
6.3.3	Slice Requirments	126
6.4	Operators Revenue and Auction Model	127
6.4.1	Price Competition: Bertrand model	127
6.4.2	Auction Mechanism for Network Slicing	130
6.5	Problem Formulation	133
6.5.1	Problem Formulation	133
6.5.2	Resource Slicing Mechanism	135
6.5.2.1	Price Auction Mechanism	135
6.5.2.2	Resource auction Mechanism	135
6.5.2.3	Optimisation Mechanism	136
6.6	Performance Investigation	138
6.6.1	Simulation Scenarios	139
6.6.2	Results Analysis	140
6.7	Conclusion	146
7	Conclusions and Future Works	147
7.1	Conclusions	147
7.2	Future Works	148

List of Figures

2.1	Device Centric Mobile Architecture	23
2.2	Cloud Central Controller Architecture	26
2.3	Hybrid Mobile Architecture: Network Centric Device Assisted	28
2.4	Reinforcement Learning Architecture	33
2.5	Network Resource Virtualization	40
3.1	Network Scenario	51
3.2	ANDSF Architecture	52
3.3	Flow Chart of Communications	53
3.4	Flow Chart of Network Selection	55
3.5	Total Average Throughput for UEs	66
3.6	Total Average Delay for Mobile Devices	66
3.7	Balance of loads between different BSs/APs in the DC sce- nario (scenario one) at selected time stamps	67
3.8	Balance of loads between different BSs/APs in the DC/NA scenario (scenario two) at selected time stamps	68
3.9	Balance of loads between different BSs/APs in the DC/NAA scenario (scenario three) at selected time stamps	68
3.10	Average Cumulative Received Throughput	71
3.11	Battery-sensitive Energy Consumption	72
3.12	Utility Values in Different Cases	73

LIST OF FIGURES

4.1	Cloud Central Controller based System Architecture	77
4.2	Communication Architecture	78
4.3	Sum throughput Vs. number of active users	89
4.4	UEs' average battery consumptions	90
4.5	Number of Handovers in Scenario One (QoS-based)	92
4.6	Number of Handovers in Scenario Two (Learning-based)	92
4.7	Number of Handovers in Scenario Three (QFM-based)	93
4.8	QoS Utility based on Equation 4.1	93
4.9	Average Jain Index Value vs. Simulation Time Steps	94
5.1	Our reference scenario with inter-slice and intra-slice priority.	99
5.2	System Architecture	101
5.3	Flow of Admission Control	105
5.4	Average Cumulative Received Throughput	113
5.5	Average Quality Levels of Different Slices	114
5.6	Time Averaged total Quality Level with Different Number of UEs	115
5.7	Percentage of free resources after the resource allocation step.	116
6.1	System model	125
6.2	Slice Model	125
6.3	Auction Flow Chart. Steps of our proposed auction model can be mainly described as: Network slices send a resource request to the slice manager which runs the auction mecha- nism considering the revenue and the network utilisation of all the chunks of the network. Afterwards, slices will run auction mechanism in order to access their optimised resources of all network chunks.	131
6.4	Cumulative Network Revenue	140

LIST OF FIGURES

6.5	Analysis of Resource Utilisation	141
6.6	Average level of satisfaction from each network chunks . . .	142
6.7	Different Requests and Different Priority for network slices for the strategy TC-NS. In case two, the 10 th network slice requests 30% more network resources, while the 20 th network slice requests 30% less network resources. In case three, the 10 th network slice holds 30% higher priority value in case three, while the 20 th network slice holds 30% less priority value. The request/priority of the 15 th network slice in three cases are the same.)	144

List of Tables

2.1	Types of Convergence for Heterogeneous Mobile Networks [1]	21
2.2	Challenges Considered in This Thesis	22
2.3	5G Characteristics Compared with 4G	35
2.4	Metrics of Virtual Networks	39
2.5	Summary of Auction Mathematical Strategies	45
3.1	Quality offered by each radio access	63
3.2	Boundary of quality level parameters depending on the appli- cation	63
3.3	Weight values of different quality level parameters and for each application type	64
4.1	Main Simulation Parameters	87
4.2	Total Number of Handovers in Each Scenarios	91
5.1	Notations in Chapter 5	110
5.2	Main Network Parameters	111
5.3	Application Slice Parameters [2]	112
6.1	Jain's Fairness Index of Each Network Chunk	143
6.2	Revenues of Selected Slices of Three Cases in TC-NS	146

List of Symbols

Symbols	Descriptions
B	Bandwidth of the Access Point
E	Energy consumption of the mobile device
P_t	Transmit power of the Access Point
σ	Received throughput of mobile users
Φ	Quality levels of mobile users
ζ	Battery discharging rate
τ	Battery capacity
ϵ	Battery consumption per distance unit
n	Propagation loss coefficient
J	The Jain's fairness index
γ	Discount rate in Q-learning
α	Learning rate in Q-learning
F	Temperature value of Simulated Annealing
κ	Temperature decrease rate of Simulated Annealing
ρ	Priority values of service slices
μ	Priority values of mobile users
η	Capacity of network chunks
ω	cost values per unit of network chunks

Chapter 1

Introduction

1.1 Introduction

With the fast development of heterogeneous mobile networks and the intelligence of mobile devices, resource providers have been asked to provide efficient radio and the core network resources in order to enhance mobile users' quality level and improve the efficiency of traffic management.

The network architecture is becoming more dynamic, distributed and heterogeneous that contains massive kinds of network technologies such as WiFi access points, LTE macrocells, picocells and femtocells, and novel radio networks. The required network traffic from mobile users is expected to be around 70% growth [3]. The new emerging network architecture will bring us new challenges of optimising traffic management by considering various quality of service (QoS) requirements and limited network resources.

Numerous researches have already modelled the future network architecture and provided mechanisms for traffic management in heterogeneous mobile networks. However, traditional methods of traffic management are not efficient in the current and the future mobile networks [4].

To address problems described above, a novel network architecture and

1.2. Research Contributions

network slicing mechanisms are proposed. These are significant elements for the future heterogeneous mobile networks. Moreover, with the ever increasing data traffic in mobile networks, an efficient traffic management mechanism needs to be designed. It should have the ability to satisfy resource requirements of mobile users and improve the efficiency of network utilisation and revenue.

In this thesis, a fully distributed traffic management mechanism is proposed, where all decisions are made at the user side. The proposed mechanism mainly focuses on mobile users' QoS requirements (e.g., throughput, delay and energy consumption, etc.) and makes fully selfish decisions. Then, the traffic management mechanism has been extended by combining a cloud central controller. The proposed central controller has been designed to address the issue of fairness among mobile users. These proposed traffic management mechanisms are used to maximise quality level of mobile users and improve the efficiency of resource allocation.

Moreover, in this thesis, novel technologies as network virtualisation and network slicing have been proposed for improving traffic management. A novel heuristic based admission control mechanism is designed. The proposed mechanism has been used to enhance the quality level of each network slice and the network utilisation. Finally, an economics based traffic management mechanism has been proposed for improving network performance from the point of network business view.

1.2 Research Contributions

The contributions of this thesis include different aspects of traffic management in heterogeneous mobile networks. Key contributions of this thesis are defining novel network architectures and designing traffic management

1.2. Research Contributions

mechanisms. These can be summarised as follows.

- A fully distributed traffic management with network analysis mechanism has been proposed. The proposed mechanism is controlled and centred by mobile users. It mainly focuses on mobile users' location profiles and multi-criteria utility values. The network analytic mechanism combines reinforcement learning algorithm which can affect traffic management decisions.
- A new efficient traffic management mechanism has been proposed. The proposed mechanism can improve quality level of mobile users while maximising fairness levels among all traffic flows. The mechanism has been designed to be maintained by a central controller on top of the whole network system. It means final decisions of traffic management are made by both mobile users and the central controller jointly.
- A novel heuristic based admission control mechanism has been proposed. The proposed mechanism considers two-tier priority level in the virtualised mobile network system. The two-tier priority order has been considered to maximise quality level of mobile users and allocate network resources efficiently. Moreover, the mechanism can dynamically allocate network resources to enabled service slices according to the current traffic load.
- A novel virtualised 5G mobile network model has been proposed to satisfy service slices' QoS requirements optimally. The proposed network model considers vertical network slicing which contains radio, computational and storage resources. Moreover, a network slice manager has been designed in this model for providing a central view of the whole network system and determining prices of different network chunks.

1.3. Publications

- A novel traffic management mechanism which combines auction and optimisation mechanisms has been proposed. The proposed novel mechanism considers both the demand and the provision in network chunks. It has been designed to maximise the network revenue and confirm efficient traffic management for optimising allocated network resources to network slices.

1.3 Publications

The publications related to the main contributions of this thesis are stated as follows.

1. (*Contribute to Chapter 3*) Menglan Jiang, "Device-Controlled Traffic Steering in Mobile Networks", The 9th International Conference on Next Generation Mobile Applications, Services and Technologies, pp.7-12, Sept. 2015.
2. (*Contribute to Chapter 4*) Menglan Jiang and Toktam Mahmoodi, "Traffic Management in 5G Mobile: Selfish Users and Fair Network", Transactions on Networks and Communications, vol.4, no.1, Feb. 2016.
3. (*Contribute to Chapter 5*) Menglan Jiang, Massimo Condoluci and Toktam Mahmoodi, "Network slicing management and prioritization in 5G mobile systems", The 22th European Wireless, pp. 1-6, May 2016.
4. (*Contribute to Chapter 3*) Menglan Jiang, Dionysis Xenakis, Salvatore Costanzo, Nikos Passas and Toktam Mahmoodi, "Radio Resource Sharing as a Service in 5G: A Software-defined Networking Approach", accept with minor correction from Elsevier Journal of computer communications, 2016.

1.4. Organisation of this Thesis

5. (*Contribute to Chapter 6*) Menglan Jiang, Massimo Condoluci, Toktam Mahmoodi and Luis Guijarro, "Economics of 5G Network Slicing: Optimal and Revenue-based allocation of radio and core resources in 5G", submitted to IEEE Transactions on Wireless Communications, 2016.

1.4 Organisation of this Thesis

This thesis is organised as follows.

- Chapter 2 gives the technical background of understanding the research area in this thesis. The chapter includes a brief summary of different aspects of traffic management in the heterogeneous mobile networks.
- Chapter 3 presents an effective traffic management mechanism for mobile users in heterogeneous mobile networks. One of the key challenges for effective traffic management is to maximise mobile users' quality level. Therefore, in this chapter, a fully distributed mechanism which mainly focuses on history records of each available network, and multi-criteria performance metrics will be described.
- Chapter 4 presents a fairness mechanism for traffic management in the heterogeneous mobile networks. High level of users' quality level and fairness among these users' flows are two important factors for researchers to consider. Therefore, in this chapter, a semi-distributed traffic management mechanism which combines selfish mobile devices and a cloud central controller will be carefully described.
- Chapter 5 presents network slicing management and prioritisation in the heterogeneous mobile networks. Network slicing, where resources are packaged and assigned in an isolated manner to set of users according to their specific requirements, is considered as a key paradigm to

1.4. Organisation of this Thesis

fulfil diversity of requirements. Network slicing has a two fold impact in terms of traffic prioritisation as it dictates for the simultaneous management of the priority among different slices and the priority among the users belong to the same slice. Therefore, a novel heuristic based admission control mechanism which considers these two tier priorities will be carefully described.

- Chapter 6 demonstrates an economics based traffic management mechanism in heterogeneous mobile networks. First, a novel network architecture will be proposed in this chapter, where network chunks are allocated to package network resources in an isolated manner to sets of service slices according to their QoS requirements (i.e., radio resource, computational resource and storage resource). Then, a novel network slicing mechanism that combines auction and optimisation mechanisms will be carefully described.
- Each contribution chapter addresses a unique research problem, and finally, the thesis is concluded in Chapter 7 with some directions for future works.

Chapter 2

Background

2.1 Introduction

Nowadays, in heterogeneous mobile networks, a major challenge is to effectively improve the quality of network connection for mobile users and meet their network resource demands. Moreover, with the tremendous increase in data traffic, another major challenge is to allocate network resources effectively to enhance network performance. In this chapter, a survey of the state of the art in heterogeneous mobile networks will be carried out. The related work on different aspects of traffic management will be summarised as well.

While 4G mobile networks are being widely deployed in the large area of the world, it has been clearly seen that unprecedented challenges in meeting users' and network operators' growing expectations will appear in the near future [5]. The exponential growth of resource requirements driven by intelligent mobile users triggered the investigation of heterogeneous mobile networks [6]. The heterogeneous mobile network is integrating (or converging) different types of access nodes (e.g., macrocells, picocells, and femtocells), and it allows each access node to share network resources with others. The purpose of integrated (or converged) networks is enlarging network capacity

2.1. Introduction

to meet bandwidth intensive resource requirements when the network resource remains limited. The integrated networks also have greater flexibility to provide higher data rate, higher quality of experience, and lower end-to-end latency and energy consumption [6] to mobile users. The integrated networks can be used to guarantee mobile users' QoS resource requirements. Intelligent mobile users in heterogeneous mobile networks have the ability to connect with more than one network access nodes using the same or different radio access technologies (RATs) simultaneously [6].

The convergence of different network technologies is one of the significant solutions that can be used to improve network performance [1]. There are four types of network convergence which are device convergence, protocol convergence, service convergence and full convergence as described in Table 2.1. Device convergence was based on integrating multiple communication interfaces within one device, and each separate communication session is controlled by the device. The protocol convergence was integrating different kinds of network protocols within a single heterogeneous network, and mobile users can handover from one protocol to the other based on their requirements. The service convergence is integrating different types of network services within the same heterogeneous mobile network. The goal of this type of convergence is keeping mobile services to work continuously, regardless of mobile devices and network handovers. The full convergence is the type planning to integrate the above three different types. The aim of the full convergence is to provide higher quality level to mobile users, even for high mobility mobile users. In order to achieve this aim, energy-efficient modulation, coding techniques, mobility management and radio resource management algorithms should be solved at the beginning.

The primary technologies of the full convergence heterogeneous mobile networks and key challenges considered in this thesis have been summarised

Table 2.1: Types of Convergence for Heterogeneous Mobile Networks [1]

Types	Objectives	Subjects	Features
Device Convergence	Convergence of different communication interfaces within one mobile device	Mobile devices and their interfaces	Simultaneous holding separate communication sessions by a single device
Protocol Convergence	Convergence of different communication protocols within one logical network	Communication protocols	Soft handover between different communication protocols without session interruption
Service Convergence	Convergence of network services within one heterogeneous mobile network architecture	Network services and mobile devices	Service has been maintained continuously, regardless of mobile device and connection type
Full Convergence	Full convergence of any mobile devices, communication protocols and network services	Network services, mobile devices, and communication protocols	Independence between mobile devices, communication networks, and service providers

2.2. Traffic Management Architectures

in Table 2.2.

Table 2.2: Challenges Considered in This Thesis

HetNets	Challenges
Mobility Management	Seamless handover between service providers, and offloading traffic resource of heavy access nodes
Traffic Management	Quality maximisation, fairness resource allocation
Heterogeneous Network Support	interference management, adaptive power control, device/location discovery
Wireless Network Virtualization	resource allocation, revenue/price optimisation, isolation and mobility management

2.2 Traffic Management Architectures

2.2.1 Distributed Device/User-Centric Architecture

In heterogeneous mobile networks, the network architecture will be changed from the base-station-centric architecture to the device-centric architecture [5, 7]. The device-centric architecture can drive mobile users to make their decisions on to which base station or access point to connect by optimizing the individual performance metric. The benefit of device-centric is in the independence of mobile users to make their decisions and the ease of taking mobile users' preference into account for maximizing their quality levels.

In order to improve quality levels, mobile users establish both down-link and uplink connections, while carrying both control and data traffic information according to where they have located [7]. The device-centric

2.2. Traffic Management Architectures

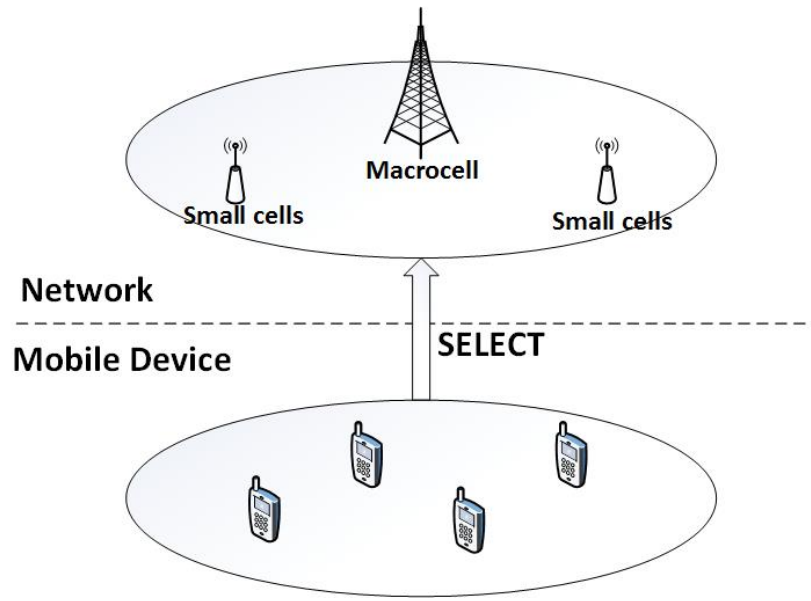


Figure 2.1: Device Centric Mobile Architecture

architecture is a fully distributed architecture, in which, mobile users are independent to make their own decisions to connect with the network they wish to. The distributed architecture is a benefit for helping mobile users to mainly consider their own individual objectives, such as network channel conditions (e.g., Signal-to-interference-plus noise ratio, SINR), information flows with different priorities and location profiles.

Generally, in a certain area of heterogeneous mobile networks, a mobile user will select to connect with the access point who can provide the highest value of quality values. The distributed device-centric architecture can be described as Fig 2.1.

Advantages of using device/user-centric architecture are as follows:

- **Efficient Information Collection:** Network densification generates more challenges in the field of efficient resource management. From the aspect of mobile users, distributed device-centric mobile architecture can receive network information efficiently based on their own location profiles and QoS requirements.

2.2. Traffic Management Architectures

- **Traffic Offloading:** Mobile devices are becoming more and more intelligent. They have abilities to implement more complicated algorithms that can help them to connect with their required access nodes, and improve their quality levels. Meanwhile, offloading traffic loads from base stations/ access points has been implemented more effectively compared with base-station-centric architecture.
- **Resource Allocation:** Network resources (e.g., CPU, storage, bandwidth, etc) needed by mobile users will inevitably generate challenges to base-station-centric architecture in the field of satisfying QoS requirements for mobile users. Device/ user-centric architecture can efficiently improve the resource allocation for mobile users.

In [8], a real-time video sharing scheme based on requirements from user-centric networking (UCN) has been proposed. The UCN can understand the context of mobile users, can profile current interests of mobile users and can personalise content delivery for mobile users [9]. A low power mobile device-centric access point congestion detection (MACoD) mechanism has been proposed in [10]. The aim of this mechanism is to help mobile devices to recognise congestion situation autonomously and take proper actions quickly in order to maintain their continuous connections with access points when the network quality becomes deteriorated. In [11], an integrated mobile device-centric approach has been proposed for data offloading. The proposed approach was focused on that a mobile device has an ability to scan conditions of alternative wireless networks, one of which they can connect with. The advantage of this approach was delay tolerant for many types of applications. Device-centric solutions for emergency communication has been proposed in [12] which is aiming to improve communication resiliency.

2.2.2 Network Assisted Architecture

With the development of network technologies, device centric network assisted architecture combined with a central controller (e.g., Software-defined Networking controller, SDN controller) has been proposed in several literatures [13, 14, 15, 16, 13, 17, 18]. The central controller can be used to decouple the function of **control** plane from the **data** plane. The control plane logically controls the operation of network switches while the data plane carrying out simple packet forwarding payload data through the physical network of switches and links. Such decoupling allows quicker provisioning and configuration of network connections [19].

The central controller can be used to abstract entire physical network resources. It can also operate network elements with intelligent strategies through an OpenFlow interface between the physical infrastructure layer and the central control layer. There is another OpenFlow interface [20] between the central controller and the applications layer [21] at the same time. The architecture of the central controller can be described in Fig 2.2. It contains three different layers: the application layer, the control layer and the infrastructure layer.

- **Application Layer:** This layer contains diversity types of software applications based on different network services. Applications can communicate with the control layer directly through an OpenFlow interface.
- **Control Layer:** This layer is the most significant part of the central architecture. The central controller logically maintains the dynamically changed network information from a global view. Moreover, the central controller can take QoS requirements from the application layer and manage the network devices via standard protocols such as OpenFlow

2.2. Traffic Management Architectures

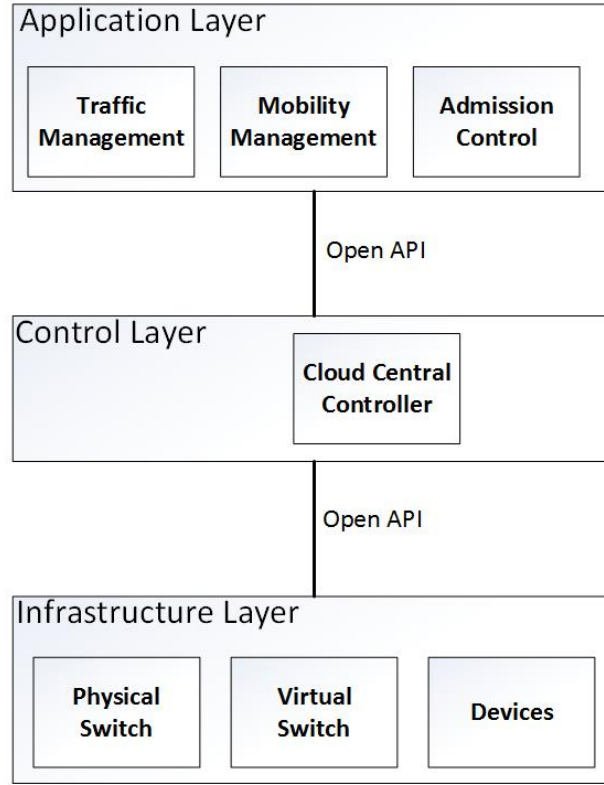


Figure 2.2: Cloud Central Controller Architecture

[20].

- **Infrastructure Layer:** This layer contains physical switches, virtual switches and massive types of mobile devices. Devices in this layer are programmable [22].

Network intelligent can be realised centrally in the central controller [23], and applications can accept instructions without understanding thousands of protocol standards.

The network assisted architecture makes network resources to be optimised by using a central controller. The architecture can offer a promising alternative for traffic management by programmatically configuring forwarding rules [22]. Moreover, the architecture allows different QoS requirements to be satisfied at the same time, and allows different kinds of services to be fulfilled with the most suitable network resources. There are three different

2.2. Traffic Management Architectures

types of network resources which are computational resource, storage resource and the radio network resource. The architecture can also effectively manage network resources globally from three aspects as follows [24]:

- **Network Selection:** Network selection has been defined to monitor and select networks according to the increased traffic demands periodically. Network selection mechanisms are proposed in this thesis (Chapter 3 and Chapter 4) where a network analytic mechanism and the central controller have been designed to improve the network performance. The latest and history information profiles of network status and mobile users have been held in order to be used in the future. Under this circumstance, consecutive requirements from mobile users can be effectively implemented by taking advantage of stored information profiles.
- **Network Handover:** In heterogeneous mobile networks, it is necessary to effectively allocate time and frequency resources within limited physical resources. Actually, mobile users are moving in a dynamic manner which will cause frequent network handovers between different access points and resource competition in the same access point. These situations will make network resource to be wasted. Therefore, in network assisted architecture, designing efficient handover algorithms can make efficient use of network resources from a global view.
- **Global Management:** Radio resource management is the most important problem in heterogeneous mobile networks. The SDN central controller in heterogeneous mobile networks acts as a global optimiser for resource managing among different types of network technologies. The aim of efficient resource management is to maximise network efficiency, and improve the quality level of mobile users with multiple

2.2. Traffic Management Architectures

interfaces. Several kinds of intelligent schemes have been proposed in literatures [24, 25] to improve the efficiency of radio resource management.

The architecture of this device-centric network assisted architecture is described in Fig 2.3. After mobile devices requesting their desired network resources, network assisted central controller will send back information of available networks for helping mobile devices to make their decisions. All the information of networks and mobile devices will be updated and stored in the central controller in real-time.

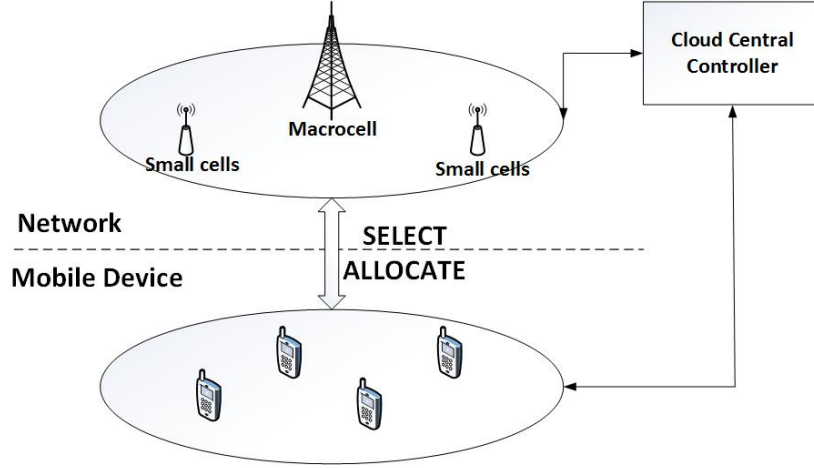


Figure 2.3: Hybrid Mobile Architecture: Network Centric Device Assisted

In [26, 14, 27], SDN based decoupling network architecture has been summarised. SDN based central controller changes the traditional network state by separating the control plane from the underlying data plane, and introducing the ability to program the network [14]. Based on this architecture, the separated control plane offers a greater control to networks through programming [28, 29], and the SDN architecture facilitates the evolution of networks[14]. In [30], an integrated architecture combining SDN and 5G has been proposed. The proposed architecture can effectively use the frequency spectrum for considerably enhancing the network performance. In [31], SDN based network virtualisation and service slicing strategies in the

next generation network have been described.

2.3 Device Centric Network Assisted Traffic Management

In heterogeneous mobile networks, multiple radio access technologies such as WiFi and LTE are coexisted [32] in order to provide more efficient network resources to mobile users according to their QoS requirements. Each mobile user has multiple interfaces and can connect with its optimised network access at any time and anywhere. One of the major issues in heterogeneous mobile networks is to maintain continuous connections for mobile users by making network selection decisions. Decisions of network selection can be made by both mobile users and the network controller. In this section, relative algorithms of device centric network selection and traffic management will be mainly introduced.

The process of network selection and traffic management can be divided into three main phases which are network discovery, network selection decision and network selection execution. In the network discovery phase, the mobile user with multiple interfaces can discover the status of the available networks and information of available network services. In the network selection decision phase, the mobile user makes a decision on which optimal network can be selected by implementing network selection algorithms. One of the most important parts of this thesis is implementing different kinds of network selection algorithms in order to improve traffic management level for mobile networks. The proposed network selection algorithms can be separated into several classes such as Fuzzy logic [33, 34], genetic algorithms, utility functions [35, 36], multiple attribute decision making (MADM) and artificial intelligent algorithms. MADM algorithms can be used to represent

2.3. Device Centric Network Assisted Traffic Management

promising solutions to mobile users with lower complexity. It can also help mobile users to select the most suitable network accesses in terms of their QoS requirements. When implementing network selection algorithms by distributed mobile users, different types of performance parameters should be considered as listed below:

- **Device side information:** QoS requirements (e.g., throughput, latency, jitter, etc), user preference, battery, etc.
- **Network side information:** network profiles, current status, etc.

Finally, in network selection execution phase, vertical handovers need to be executed in a seamless manner in order to help users to dynamically select the best network in terms of their quality of service [37].

2.3.1 Multi-Attribute Decision Making

In heterogeneous mobile networks, one of the network selection problems is the multi-criteria nature of mobile users. Therefore, Multi-Attribute Decision Making (MADM) algorithms were proposed to ensure a successful allocation of complex network resources after making a network selection decision.

Many types of research were carried on to ensure that MADM algorithms can be used to implement seamless handovers. In [32], a large number of MADM algorithms have been compared, such as Simple Additive Weighting (SAW), Technique for Order Preference by Similarity to Ideal Solution Algorithm (TOPSIS), and Analytical Hierarchical Process (AHP) etc. These algorithms have different activities and can be used to deal with network selection problems. SAW and TOPSIS are used for ranking alternative networks in terms of quality of service can be offered, while AHP is designed to assign weights to various attributes.

2.3. Device Centric Network Assisted Traffic Management

In [38], the TOPSIS based network selection model was proposed. The TOPSIS model has been used to define index values. These index values can be used to compare normalised distances values from mobile users to all alternative networks. The best solution for mobile users is the one that holds the shortest normalised distance to the best network and the longest normalised distance to the worst network. Although TOPSIS is used to rank available networks, there still exist some points of weakness in this method as listed below [39]:

- **Inefficient decisions:** Adding or removing one alternative network from the candidate list, the list of remaining candidates will be changed. It will cause inefficient network selection decisions for mobile devices.
- **Resource waste:** Unnecessary handovers will waste network resources and increase processing overheads.

Therefore, in [39], iterative TOPSIS approach which ranks only the top candidates and removes the lowest candidate after each iteration has been proposed to solve these problems. However, though mobile users can seamless connect with candidates in this literature, performance definition of multiple criteria and quality levels of each mobile devices have not been considered.

In [40], a SAW based network selection mechanism was modelled and applied to provide a ranking order of all alternative networks. By using SAW, the overall score of alternatives was calculated by the weighted sum of all attribute values [41]. In [42], a fuzzy handover decision strategy was introduced by considering QoS handover criteria. The fuzzy logic has been applied to deal with the imprecise information of some criteria and user preference. In [42], the fuzzy MADM algorithm was analysed to be a feasible approach. The approach of fuzzy logic based handover initiation and the decision has been proposed in [43] as well. The approach converted performance values

2.3. Device Centric Network Assisted Traffic Management

of network alternatives to fuzzy numbers and then made decisions based on heuristic decision rules. In [44], an enhanced vertical handover technique which combines two MADM methods had been proposed. The enhanced technique combined the analytic network process (ANP) and TOPSIS together in order to satisfy Always Best Connected (ABC) requirements. In [45], a combinatorial network selection algorithm which combined AHP and Gray Relational Analysis (GRA) was proposed. After weighted by AHP, the algorithm GRA was applied to rank alternatives as well. AHP can decompose the network selection problem into several sub-problems and weight values are assigned to each sub-problem [41]. Moreover, in [46], network selection based on AHP and TOPSIS was proposed. The AHP method was used to calculate weight values of different criteria while TOPSIS was used to determine the ranking order of alternatives.

2.3.2 Reinforcement Learning

Reinforcement learning (RL) is an important approach for improving the efficiency of network selection and traffic management based on the history experiences in heterogeneous mobile networks. RL is one of the machine learning approaches which is concerned with how an intelligent agent can take actions in an unknown environment in order to maximise their cumulative reward values [47]. Machine learning approaches can help networks to achieve context awareness and intelligence for enhancing network performance [48]. The basic architecture of reinforcement learning can be expressed as Fig 2.4. The intelligent agent in the RL architecture can make actions in order to affect the current state of the environment through interacting with each other continuously [47]. The reward value provided by the environment can help intelligent agents to make better decisions.

Reinforcement learning has two characteristics [49] as listed below:

2.3. Device Centric Network Assisted Traffic Management

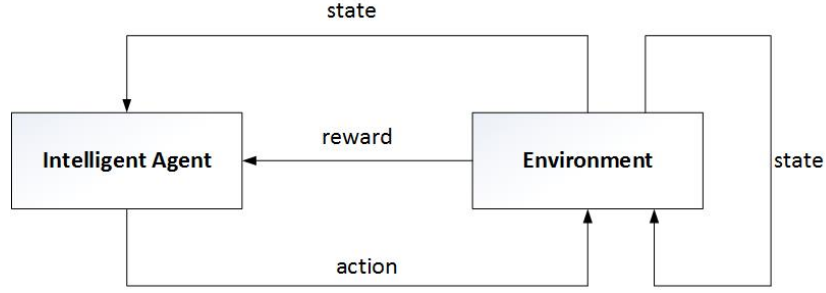


Figure 2.4: Reinforcement Learning Architecture

- RL represents the performance metrics as a whole without modelling the complex operating environment based on each single factor. For instance, instead of tackling every single factor that affects network performance such as wireless channel condition or mobility, RL monitors the reward resulting from a whole view point. These reward values cover different types of user preferences and can help mobile users to select optimal actions.
- RL does not build explicit models of the other agents' strategies or preferences on action selection.

Moreover, there are two different approaches that can help intelligent agents to find optimal actions which are model-based approach and model-free approach [47]. The model-based approach requires known reward values and transition functions while the model-free approach considers unknown reward values and transition functions. Q-learning is one of the popular model-free reinforcement learning approaches and has been widely applied in wireless mobile networks. Components of Q-learning can be divided into three parts which are:

- **State:** A set of states represent the decision making status determined by local network environment. At any time step t , the alternative network i has its own state S_i^t .

2.3. Device Centric Network Assisted Traffic Management

- **Action:** A set of available actions represent actions such as *transmit*, *suspend* or *stop* based on the information of local network environment which is recorded by Q-values. In terms of continuous records of Q-value, an action of the alternative network i at time t , a_i^t , represents state transmission from one state of a network at time t , S_i^t , to another state of a network at time $t + 1$, S_i^{t+1} .
- **Reward:** Reward values represent the effect of actions. Whenever an action a_i^t has been carried out, a reward value R_i^t will be generated for network selections. The reward value can be evaluated in terms of some performance metrics, such as throughput, transmission delay, energy consumption and channel congestion level [48].

In [49], RL was applied to achieve context awareness and intelligence in wireless networks. The context awareness and intelligence enable to observe, learn, and response dynamically changed operating environment in an efficient manner. Meanwhile, the RL approach can be used to provide network performance enhancement. In [47], a cooperative multi-agent reinforcement learning framework towards multiple access control (MAC) in wireless sensor networks has been proposed and open challenges related in employing RL in wireless sensor networks had been concluded as well. In [48], RL has been proposed to address routing challenges by enabling wireless nodes to observe and gather information from their dynamic local operating environment and RL has also been proposed to make efficient routing decisions.

2.4 Novel Traffic Management in the 5G Mobile Networks

2.4.1 The Future Network Technology: 5G

5G mobile networks provide much higher data rates, lower latency, higher network capacity, higher spectral efficiency, and higher users' quality of experience, compared with those in the 4G mobile networks [3] as described in Table 2.3. The architecture of 5G mobile networks has been explained in [50] and [51], and the network capacity has been considered as a main characteristic for evolution compared with previous network generations. The increased network capacity can satisfy the higher traffic demands from mobile users, and it is generated by a large number of low power small cells which serve for data only services [52]. Moreover, with the help of massive MIMO technology, energy efficiency, spectral efficiency, and data rate of the network have been improved as proposed in [53, 54].

Table 2.3: 5G Characteristics Compared with 4G

Characteristic	Values (Compared with 4G)
Network Capacity	about 100 times
Spectral Efficiency	about 10 times
Energy Efficiency	about 10 times
Battery Life Time	about 10 times longer
Latency	about 5 times shorter
Number of Connecting Devices	about 10 – 100 times
Reduction in energy per bit	about 100 times

Several possible research issues raised by the 5G mobile network [3] have

2.4. Novel Traffic Management in the 5G Mobile Networks

been listed as follows.

- **Network Densification:** In heterogeneous mobile networks, different types of network technologies have been converged. The density of small cells and wireless access points is expected to be higher for satisfying network traffic requirements.
- **Central Cloud Architecture:** The mechanism of the cloud central controller in the heterogeneous mobile networks, such as SDN controller, can be used to offer cost effective and energy efficient solutions for improving network performance. Though many researchers are working on the topic of building central cloud architecture, it is still more challenging for us to design a more efficient controller mechanism combined with the 5G mobile networks.
- **Low Latency and High Quality Level:** The round trip latency has been identified as 1ms [55] in 5G mobile networks. Except for higher data rate, lower latency is crucial for helping mobile users to achieve high quality level. With the fast development of the Internet and intelligent mobile devices, there are various kinds of QoS requirements from different service types. Therefore, improving quality levels for mobile users has presented a number of research challenges due to its subjective nature.
- **Energy Efficiency:** Cost and energy consumption are the other major considerations for mobile users in the 5G mobile networks. Despite a variety of works have been done in the field of energy efficiency, it still offers a huge scope of improvements, especially under the 5G mobile networks.

2.4.2 Network Virtualization in 5G

Network virtualization has been considered as one of the most significant technologies for the future networks [56], [57]. By allowing different kinds of network technologies to coexist on a shared physical infrastructure and maximising network utilisation, network virtualization provides flexibility of resource allocation, promotes diversity of resource types, and promises increased manageability of resource management [56].

Network virtualization can abstract, slice, isolate and share wireless physical infrastructure and network resources into virtual resources with holding certain corresponding functionalities [58]. Since wireless network resources are sliced into multiple slices, the terms of virtual slice and virtual network have the similar meanings of virtual wireless network resources [58]. Network virtualization is defined by two independent entities which are infrastructure providers (InPs) and service providers (SPs) [56].

- **InP:** Manage physical infrastructure resources and provide these network resources to different SPs through programmable interfaces.
- **SP:** Create virtual networks (VNs) by aggregating resources from multiple InPs and program allocated virtual resources. SPs can also provide end-to-end services to end users who require, and provide network services to other SPs [56].

In order to materialise network virtualisation, different kinds of requirements have been described as follows:

Coexistence: The heterogeneous mobile network contains different kinds of network technologies, the purpose of network virtualisation is to allow these network resources to be shared by different mobile users. Moreover,

2.4. Novel Traffic Management in the 5G Mobile Networks

virtual network resources have been sliced based on the requirements of different SPs, and each slice is different with the others. Therefore, network virtualization has to have the ability to integrate different virtual slices who hold various QoS requirements, topology, services type, security level and user behaviours [59].

Flexibility: Network virtualization provides freedom in different aspects of networking by decoupling customised control protocols from the underlying physical networks and other coexisting VNs [58]. The flexibility depends on the level of virtualisation, the higher level of virtualisation the lower flexibility.

Manageability and programmability: Virtual network resources have been sliced and assigned to SPs which is decoupled from InPs. Network virtualisation provides complete end-to-end control of virtual resources to SPs, while SPs have the ability to manage configuration and allocation of these virtual network resources. Programmability is another indispensable requirement which has been integrated into network virtualization to ensure flexibility and manageability. Programmability can also help SPs to implement customised diverse services, protocols and networks. Programmability needs network operators to provide appropriate interfaces, programming language and a secure programming paradigm with a considerable level of flexibility [60].

Isolation: Because of the high interference between different network cells, isolation becomes a more significant and complicated requirements in heterogeneous mobile networks. Isolation requires any change in one virtual resource slice should not generate changes in other service slices [61]. It also means that one virtual resource slice does not know the existence and the

2.4. Novel Traffic Management in the 5G Mobile Networks

Table 2.4: Metrics of Virtual Networks

Metrics	Description
Throughput	achieved data rate from selected virtual nodes;
Delay	the time a packet needs to be transmitted from the source node to the destination node;
Jitter	the variance in inter-arrival times of packets from the source node to the destination node;
Cost	the sum of all spent for allocating network resources;
Revenue	the sum of all receiving by providing network resources.

status of other virtual slices.

Stability and convergence: Though isolation ensures that status in one virtual resource does not affect status in other virtual resources, stability can decrease the effects of errors and misconfiguration in the underlying physical networks. Moreover, convergence allows virtual networks to converge to their stable status in case of any instability.

Resource utilisation: Wireless network virtualisation has been proposed to guarantee the efficiency of network resources (e.g., radio resources, computing resources, and storage resources, etc) utilisation. Performance metrics of network virtualisation should also be efficiently designed in order to improve the network performance. Two types of performance metrics have been widely considered which are QoS-based metrics and cost-related metrics. QoS-based metrics aim to measure the impact of different service requirements when they are planning to use virtual network resources. The cost-related metrics aim to maximise the network revenue. Performance metrics of network virtualisation have been described in Table 2.4.

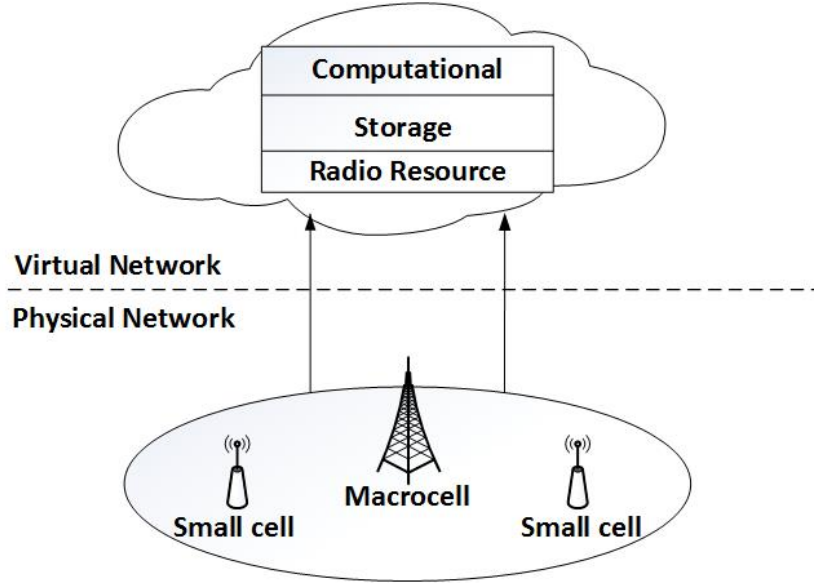


Figure 2.5: Network Resource Virtualization

2.4.3 Network Slicing in the 5G Mobile Networks

Wireless network virtualization [62, 63] is an important research direction in heterogeneous mobile networks because of the concept **slicing**. The concept of network slicing has initially been proposed for the 5G mobile networks [64]. A 5G slice is a collection of logical network functions that depends on supporting different types of service requirements [65]. Each slice is considered as a combination of network resources (e.g., radio network resources, the core network resource and the edge network resource) and operates on top of the physical infrastructures with virtual network performance [66], it has been described in Fig 2.5. Moreover, each slice has a dedicated treatment according to the service requirements. Service requirements in the 5G mobile networks become more and more diverse, while all different types of service requirements will not be satisfied at the same time via one uniform network architecture. Therefore, the idea of network **slicing** has been created for a variety of services over a common network infrastructure [65].

Based on different types of service requirements from mobile users, vir-

2.4. Novel Traffic Management in the 5G Mobile Networks

tual network resources have been implied three different slicing methods as described below:

- **Spectrum slicing** has been considered as an extension of dynamic spectrum access and spectrum sharing. In wireless network virtualisation, spectrum slicing is data slicing instead of physical layer technology.
- **Infrastructure slicing** has been considered to share virtualised physical network elements (e.g., base stations, processor hardware, and routers, etc.) among multiple operators. In wireless network virtualisation, infrastructure slicing can be used to reduce network congestion and improve network resource utilisation.
- **Network slicing** has been considered as the combination of spectrum slicing and infrastructure slicing. An access node in heterogeneous mobile networks can be virtualised into multiple virtual access nodes, while the virtualised network resource of the whole network system (e.g., radio resource, computational resource and storage resource) has been sliced and assigned to mobile users.

From the network aspect, different kinds of slice creating technologies have been proposed in [67, 68] such as link layer slices and switch slices. In wireless network virtualisation, traffic management is the most significant challenge that should be taken into account. Admission control and auction mechanism are two important techniques for effectively allocating network resources and enhancing network revenue.

2.4.4 Admission Control in Virtualised Mobile Networks

In network virtualization, admission control has been proposed to maximise the utilisation (revenue) while guaranteeing the QoS of existing mobile users by controlling the admission of incoming users. Admission control can confirm allocated virtual resources do not exceed the total available capacity of physical networks [58].

The strategy of admission control is proposed to allocate network resources to mobile users more reasonably. At the initial stage of the strategy, mobile users access network nodes based on their resource requirements. Considering congestion and traffic offloading of the mobile network, virtual network resources should be sliced into multiple resource slices from the aspect of frequency domain and the time domain. Resource slices are minimum available units for mobile users and each resource slice can only be assigned to one mobile user. Then, the sliced resource of network access nodes will be redistributed to mobile users. The admission control strategy at this stage considers both QoS requirements of mobile users and available network resource to admit or reject the new mobile user entering the network for the purpose of traffic management of the current mobile network. The admission control strategy can improve the communication reliability, reduce the communication congestion and enhance the resource utilisation of the network. Finally, mobile users have been reallocated network resources based on decisions made by admission control strategy.

In [69], a new LTE mechanism for allocating network resources in the downlink based on call admission control has been proposed. The proposed mechanism took into account mobility within and between cellular users. It has been proposed to accept each call by maintaining its initial throughput

2.4. Novel Traffic Management in the 5G Mobile Networks

regardless of the position of the mobile user in the network. In [70], a joint resource provisioning and admission control policy have been proposed for an orthogonal frequency division multiple access (OFDMA)-based wireless virtualised networks. The proposed policy has been used to deal with the problem of stochastic nature of channels between mobile users and access node and limited available network resources. The proposed policy has been evaluated to maximise the total rate of the network while adjusting minimum QoS requirements. These proposed admission control mechanisms can be extended to the 5G mobile networks, but with the limitation of less considering different types of services requirements from the business point of view.

An agile admission control mechanism which is based on predictions of users' quality level has been proposed in [71]. The mechanism was built based on Markov Decision Process (MDP) and aimed to derive the optimal policy according to the traffic load for maximising users' quality level. In [72], an admission control based handover mechanism was proposed in high dense femtocell networks. The admission decisions of network handovers have been made for new calls and current calls. In [73] and [74], the admission control mechanism has been formulated as a Semi Markov Decision Process (SMDP) in order to minimise the sum of overall blocking probability. Though the proposed admission control mechanisms have been widely used in optimizing mobile users' quality levels, they were less considering efficient network utilization from the global point of view.

In [75] and [76], joint power and admission control (JPAC) algorithms have been proposed to solve the problem of maximising the number of mobile users in single tier networks. In [77], two efficient JPAC algorithms in single and multi-tier underlay cellular networks have been proposed. The algorithms have been evaluated to be low complexity and the performance

2.4. Novel Traffic Management in the 5G Mobile Networks

of these algorithms are superior to that of existing ones in terms of average outage ratio of low-priority users. In [78] and [79], several centralised and distributed JPAC algorithms have been proposed for two-tier cognitive radio networks in order to maximise the network resource utilisation. The limitation of these proposed admission control approaches is that they were less considering priority levels of different network slices and the priority levels of mobile devices belong to the same slice.

2.4.5 Auction Model in Virtualised Mobile Networks

In virtualised mobile networks, an auction-based business mobile network mechanism has been considered as one of the technologies for network traffic management [4]. The auction mechanism can dynamically and effectively allocate network resources among mobile users and network providers. The components of auction mechanism in virtualised mobile networks can be divided into three parts which are buyers, sellers and auctioneers [80]. Buyers are mobile users who are planning to pay for receiving network resources based on their QoS requirements. Sellers own network resources (e.g., bandwidth, licenses of the spectrum, and time slot [4], etc.) and earn revenue by selling resources. Auctioneers are central cloud controllers acting as intermediate agents which can control and conduct auction processes.

The aim of auctioneers is improving the utilisation of virtualised network resources while satisfying QoS requirements from mobile users. Therefore, auction-based traffic management algorithms are needed to assign virtualized network resources to service slices. The main metrics of auction-based algorithms are network **revenue** received by each virtual network and **cost** paid by each virtual network [21]. The revenue received by each virtual network indicates the gain of the virtual network, while the cost per virtual network indicates the physical resources that are expended to accept a vir-

2.4. Novel Traffic Management in the 5G Mobile Networks

Table 2.5: Summary of Auction Mathematical Strategies

Strategies	Features	Scenarios
English	Open-cry bidding process; Bidding price monotonically ascending;	Sellers maximize revenue
Dutch	Open-cry bidding process; Available bidding prices quoted by seller; Bidding price monotonically descending;	Auctioning perishable commodities
VCG	Generalized Vickrey Auction; Truthfulness guaranteed;	Widely applied
Combinational	Multiple auction commodities; Some commodities are valued most;	Allocate resources as CPU, etc

tual network [57]. The revenue-to-cost ratio provides a metric relating to both revenue and cost and a higher ratio value means a more optimal traffic management.

The auction mechanism contains different kinds of mathematical strategies and has been summarised in Table 2.5 [4]. The auction mechanisms can be used for multi-player optimisation. Moreover, the auction mechanism can be used to improve traffic management by achieving individual optimal solutions of mobile users and optimal overall traffic revenue of networks.

Game theory [81] is a typical mathematical tool to analyse behaviours of buyers and sellers in auction processes. It has been extensively used to address wireless network problems such as resource and interference management. In [82], a distributed game theory mechanism has been proposed for relay selection and power control in multi-user cooperative communication networks. By using this mechanism, the utility of relay nodes has been maximised. In [83], a game theory based channel allocation scheme has been

2.4. Novel Traffic Management in the 5G Mobile Networks

proposed in cognitive radio networks. The proposed scheme was used to capture the selfish and the cooperative behaviours of access nodes. In [84], a Stackelberg game based resource allocation mechanism has been proposed for maximising network utilisation. The Stackelberg equilibrium has been effectively used in distributed power allocation and minimising network overhead for spectrum sharing networks. Moreover, in [85] and [86], evolutionary game theory (EGT) based distributed resource allocation schemes have been proposed for downlink transmission in OFDMA-based small cell networks. The EGT mechanism can be used to address the problem of resource allocation and provide fairness of allocated resources among mobile users.

Chapter 3

Distributed Traffic Management in the Heterogeneous Mobile Networks

3.1 Introduction

Today's mobile devices, including smart phones and tablets, have strong computational power and multiple radio interfaces of different communication technologies, including Wi-Fi and cellular. In order to address the problem of growing traffic demands driven by mobile users, service providers design dense networks (i.e. cellular base stations and Wi-Fi access points) and use various techniques (e.g. offloading) to effectively allocate users' traffic to radio access networks [87, 88]. An efficient utilisation of overall available network resources can improve the quality level of mobile users. Traffic management aims at distributing the network resource in an optimal manner across different radio access networks. However, several challenges have

3.1. Introduction

been introduced in wireless mobile networks, such as significant interference, frequent handover, and complex back-hauling.

Mobile devices are the best points in the network to make certain decisions on the management of radio and network resources. To this end, a fully distributed traffic management approach has been proposed, controlled and centred by intelligent mobile devices, i.e. the User Equipments (UEs), which can make final decisions on where to direct their traffic by taking advantage of the analytics of the network. It can be foreseen that such a device-centric and device-controlled networking has a more radical move in flattening the mobile network architecture and in re-layering the integration of mobile core with the Internet. Given their diversity in hardware capability and service requirements, mobile devices are in the unique position of identifying the radio coverage, their platform characteristics, and applications' requirements. Furthermore, mobile devices can gather network information through the 3GPP standard S14 interface and through the network analytics of the Access Network Discovery and Selection Function (ANDSF) server [89]. ANDSF [90] is defined in 3GPP in order to support traffic offload with universal connectivity [91]. ANDSF allows UEs to discover available access networks based on their positions and to select one which can best satisfy their requirements among the access networks.

In this chapter, two kinds of different network selection mechanisms have been modelled by using Multi-Attribute Decision Making (MADM) mechanism combined with network analytic technology. A multi-criteria utility function has also been defined based on RSS and battery consumption combined with the Reinforcement Learning (RL) mechanism. The RL mechanism can analyse network information to reinforce the effect on the traffic management decisions for UEs. The network selection algorithms aim to maximise the user-based Quality of Service (QoS), that is the quality of

3.2. Related Work

users' experience. Meanwhile, the benefit of using analytic-based selection algorithm is the potential reduction in the frequency of handover, since the selected AP/BS could remain the optimal choice for a longer period of time.

The remainder of this chapter is organised as follows. Section 3.2 briefly reviews the state of the art in using RL for selection of radio access, the benefits of using ANDSF analytic server and the using the multi-criteria utility function. After elaborating the system model in Section 3.3, the proposed utility functions and selection algorithms will be described in Section 3.4. Section 3.5 presents simulation study and performance observations. Finally, a summary of this work and road ahead is detailed in Section 3.6.

3.2 Related Work

In this section, the state of the art in traffic management and selection of radio access nodes in heterogeneous mobile networks have been reviewed as well as those research works that are focused on the device-centric and device-controlled networking.

The proposed technique in [92], design two entities at both the UE side and in the core network. The entity at the UE side prioritise available RANs based on the requirements of running applications, and the final decisions for RAN selection, are made at the core network. The network selection technique in [93], deploys the multi-attribute decision making (MADM) algorithm. In this work, a classification method was applied to build classes which having the similar criteria, and then Analytic Hierarchy Process (AHP) method was applied to determine weights of different classes.

Furthermore, various approaches for managing traffic in heterogeneous network is discussed in [94], where fundamental concepts in the design of generic traffic management algorithms are presented. In this white paper,

3.2. Related Work

traffic management algorithms are classified based on whether load, coverage, policy or a combination of those decides on distributing traffic. It has been suggested that combining these factors, and designing a more sophisticated traffic management algorithm, can result in better system-level performance.

The device-centric network architecture has been discussed in Chapter 2, as a solution that can address users' stringent QoS requirements in the next generation mobile networks. Both spectral and energy efficiencies have been demonstrated within the device-centric architecture in [87]. In [95], automatic Access Network Selection (ANS) mechanism has been proposed in a device-controlled manner. This scheme takes into account multi-criteria such as the quality of the connections, the preferences of the end users and the cost, in order to help users to select the best access networks. In [96], critical aspects and research challenges of ANS are discussed, while the proposed ANS is device-controlled. In this chapter, the network analytic mechanism has been added in the system model. Because it can help UEs to consider network information when they decide where to connect with.

The 3GPP standard introduced ANDSF can support radio access information of different access networks [97, 89]. According to the standard, ANDSF server acts as a standalone entity between UEs and radio access networks (RANs). It can extend policy control beyond the core network and out to UEs through its interface (S14).

Based on the previous researches and standard progresses, a device-controlled mechanism has been proposed in this chapter by using different kind of selection algorithms which are Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) and Q-learning for selecting RANs (the specific access point) while implementing traffic management by using the policy and available analytics through network analytic mechanism.

3.3. System Model

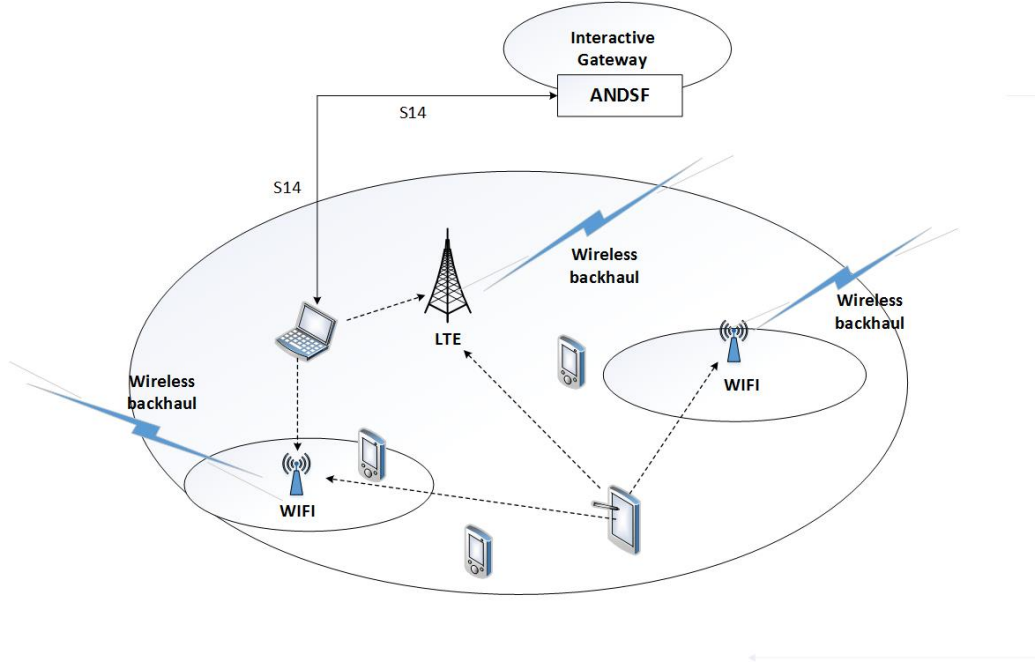


Figure 3.1: Network Scenario

3.3 System Model

3.3.1 Model of the System-level Architecture

In this section, elements of the proposed traffic management model is explained in detail, and within the network scenario of Figure 3.1. The three main elements in this model are the mobile device (i.e. the UE), the mobile core/access network, and the analytic server (i.e. the ANDSF server).

As final decisions will be made by UEs themselves, information of both wireless environment and the mobile access/core network will be assisted by the network analytic mechanism. In this network model, UEs communicate with the ANDSF server via the S14 interface [98].

Network part provides connectivity to UEs. The access network model includes four different access points, denoted by $N_i (i = 1, \dots, 4)$, consisting of LTE macro ($i = 1$), pico ($i = 2$) and femto ($i = 3$) as well as one WiFi access point ($i = 4$). The coverage area of these wireless access points is a

3.3. System Model

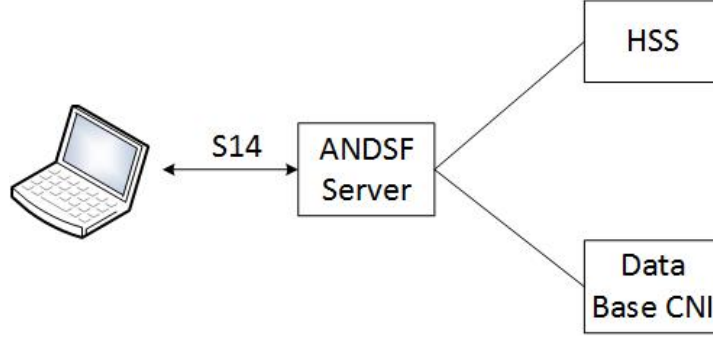


Figure 3.2: ANDSF Architecture

circle with diameter \mathbb{R}_i meter, where $\mathbb{R} = \{500, 300, 50, 100\}$ (values from [99, 100]). The Wi-Fi access points (AP) and cellular base stations are both referred to as wireless AP from this point on.

ANDSF server provides network policy and analytics. Therefore, the performance of different RANs and QoS of UEs on a per access point and per user basis is available. The information can be collected from data base Candidate Networks Information (CNI) which connects with ANDSF directly as Fig 3.2.

The performance metrics have been considered to use are channel quality and mobile battery lifetime for each UE towards alternative RANs. UEs have been defined as uniformly distributed at each time steps and they are supposed to move with the same probability to all directions. The flow chart of communication has been described in Fig 3.3.

The wireless channel is modelled with path loss and the RSS can be explained as $RSS_i = Pt_i - PL(d_i)$, where Pt_i denotes the transmit power of the i^{th} AP, d_i is the distance between UE and the i^{th} AP and $PL(d_i)$ is path loss value [101].

3.3.2 UE's Battery Models

The UEs' battery consumption and how it will be affected by the application's throughput, is modelled in this section. The model described in [102]

3.3. System Model

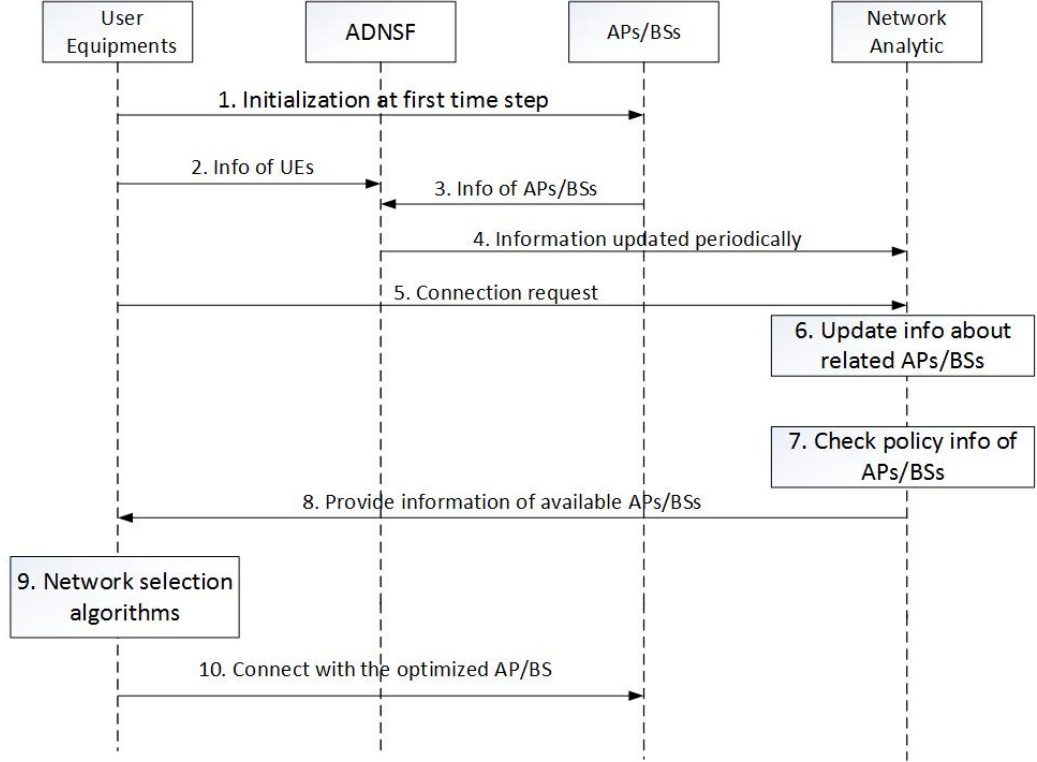


Figure 3.3: Flow Chart of Communications

has been used for the battery discharging rate ζ . During the lifetime of battery, T_l , ζ can be defined based on the Equation (3.1).

$$\zeta(\tau) = \frac{\pi^2}{3\tau^2} e^{-\tau^2 T_l} \quad (3.1)$$

where τ is the value of battery capacity and may vary from battery to battery in the range of (0.4, 1). It has been shown in [103] that when running three different applications concurrently on various smart phones, the battery lasts for two hours $T_l = 120$ mins working time.

The energy consumption of the device, when connected to the i^{th} AP, with distance d is detailed in Equation (3.2) [103].

$$E_i = \zeta(\tau) + \epsilon d_i^n \quad (3.2)$$

where ϵ and n denote the battery consumption per distance unit, and the

3.4. Device Controlled Network Selection

propagation loss coefficient consecutively.

The effect of throughput and RSS on energy consumption can also be captured in closed form [104], as follows,

$$E_i = T_i * (\sigma_i r_b + r_e) \quad (3.3)$$

where σ_i denotes UE's received throughput from the i^{th} AP. The r_b and r_e represent energy cost for transferring per bit and energy cost for connection establishing respectively. Combining (3.2) and (3.3), throughput, σ , and RSS, can be represented as a function of battery discharging rate in Equations (3.4) and ((3.5) consecutively.

$$\sigma_i = \frac{\zeta(\tau) + \epsilon d_i^n}{T_i * r_d} - \frac{r_e}{r_b} \quad (3.4)$$

$$RSS_i = (2^{\frac{\sigma_i}{B_i}} - 1) * P_i(noise) \quad (3.5)$$

where B_i is the value of bandwidth provided by the i^{th} AP. It can be indicated that if the received throughput of UE is higher, battery consumption will be higher as well. It is also clear that lower RSS will result in higher energy consumption at the device.

3.4 Device Controlled Network Selection

In this section, the proposed MADM algorithm based on network analytic mechanism has been described in detail, and then, the utility function has been defined based on the available metrics at the device side for reinforcement learning mechanism.

In order to help the UE to connect with any optimal BS/AP, the following steps will be executed when making decisions by the UE as Fig 3.4:

3.4. Device Controlled Network Selection

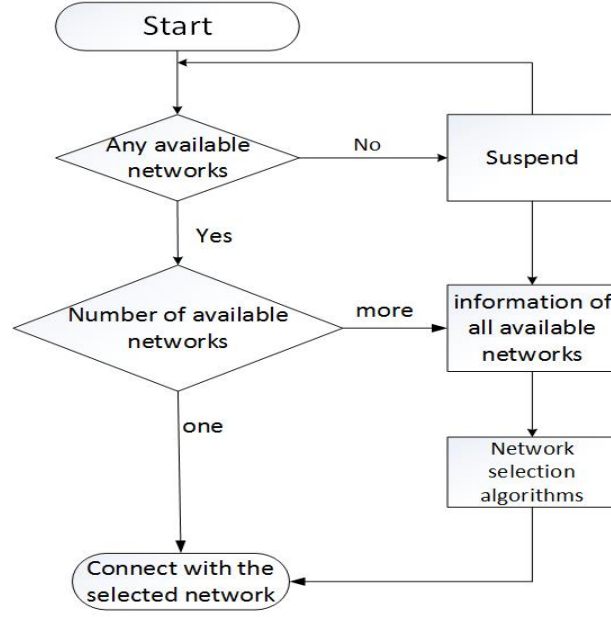


Figure 3.4: Flow Chart of Network Selection

1. If there is no APs/BSs which can guarantee the throughput required by the UE, the connection request will be suspended;
2. If only one available BS/AP can provide required throughput by the UE, the UE will connect to that BS/AP;
3. If there are more than one available BSs/APs which can satisfy the UE's throughput requirement, the UE will select the optimal BS/AP depending on the value of preference order or the value of utility functions.

3.4.1 Multi-Attribute Decision Making Mechanism

3.4.1.1 TOPSIS

From the aspect of optimising resources allocated to UEs, the selection of BSs/APs based on performance measurements and the network information is modelled by using MADM algorithms for UEs. In this implementation, *Analytic Hierarchy Process* (AHP) has been used to calculate weight values

3.4. Device Controlled Network Selection

for each selection criteria, and *Technique for Order Preference by Similarly to Ideal Solution* (TOPSIS) has been used for selecting the optimal BS/AP. The combination of AHP and TOPSIS has previously been used for network selection in [39]. TOPSIS selects the best alternative (in this case, e.g. the best BS/AP) based on multiple criteria, as the solution. Finding the best alternative is based on finding the candidate with the shortest distance from the best possible solution, while having the longest distance from the worst possible alternative. The implementation in this chapter follows six steps as below.

1. *Constructing the decision matrix for TOPSIS*: The decision matrix is $C = [c_{ik}]$, where c_{ik} represents rating of the Alternative N_i (certain access point) with respect to delivering attribute k (certain QoS parameter, $k = \{1, 2, \dots, K\}$).
2. *Constructing weighted normalized decision matrix*: the decision matrix is normalized, it will be normalized as $V_{norm} = [v_{ik}]$, such that,

$$v_{ik} = \frac{c_{ik}}{\sqrt{\sum_{i=1}^I c_{ik}^2}}, \quad (3.6)$$

where I is the total number of alternatives (in our case $I = 4$). We then calculate the weighted normalized decision matrix, $D = [d_{ik}]$, can be shaped by multiplying v_{ik} by the W_k as computed in AHP, and detailed in Table 3.3.

3. *Determining the negative and positive ideal solutions*: the negative and positive ideal solutions, i.e. A^- and A^+ , can be described as Equations (3.7) and (3.8). For criteria k that we aim for a higher value (e.g. throughput), we can write these values as:

3.4. Device Controlled Network Selection

$$\begin{aligned} A^+(k) &= \max_i(d_{ik}) \\ A^-(k) &= \min_i(d_{ik}). \end{aligned} \tag{3.7}$$

And for value of undesired criteria k such as delay and jitter, we have:

$$\begin{aligned} A^+(k) &= \min_i(d_{ik}), \\ A^-(k) &= \max_i(d_{ik}). \end{aligned} \tag{3.8}$$

4. *Calculating Euclidean distances to positive solutions and negative solutions:*

In step 4, Euclidean distances of alternative N_i to the positive and negative ideal solutions are computed as below:

$$\begin{aligned} S_i^+ &= \sqrt{\sum_{k=1}^K (A^+(k) - d_{ik})^2}, \\ S_i^- &= \sqrt{\sum_{k=1}^K (A^-(k) - d_{ik})^2}. \end{aligned} \tag{3.9}$$

5. *Ranking the preference order:*

Finally, compare results according to

$$T_i = \frac{S_i^-}{S_i^- + S_i^+} \tag{3.10}$$

and then choose the maximum value of T_i as the optimal result.

6. *Ranking abnormalities:*

TOPSIS algorithm is known to be suffered from ranking abnormalities, because if we change the number or the value of criteria, the value of normalised decision matrix will be changed at the beginning. Then, the value of positive and negative solutions will be changed as well which will affect final decisions for UEs. That means, in order to help UEs to select optimal networks based on their requirements, the ranking order

3.4. Device Controlled Network Selection

T_i should be changed in terms of changing values of each criteria at different time steps.

3.4.1.2 Network Analytics

As mentioned earlier, in addition to the network status and radio measurements, the best BS/AP could be selected based on network analytics. It can be argued that location profile is a valuable information in deciding which network can carry the traffic better. In a nutshell, if it has been known what was the experience of previous users standing in the current location of the device, and which radio access they were connected to, it can potentially make a decision only based on that.

Let for device l the location profile be P_l in which, each row represents information of one specific location, i.e. current location is in the first column (denoted by Δ), and then pair values of goodput (goodput is considered as the combination values of both throughput and delay) and BS/AP identification, denoted by (g, η) as in equation (3.11) and equation (3.12).

$$P_l^{m,1} = \Delta_m, m = 1, 2, \dots, M \quad (3.11)$$

$$P_l^{m,t} = (g, \eta)_{m,t}, m = 1, 2, \dots, M; t = 1, 2, \dots, T \quad (3.12)$$

where M and T identifies a total number of location profile has been stored. For simplicity of communication between mobile devices and networks, it has been assumed that the network can send only two anonymous rows from the profile to the mobile device. Those two rows correspond to the BS/AP that offered best and worst experience in the then meter vicinity of the device. Then, these two values can be used in equation (3.9) and instead of the A^+ and A^- . Therefore, it can be found that the solution is

3.4. Device Controlled Network Selection

based on having furthest distance from the worst experience recorded in the profile, and having the closest distance to the best past experience. Using such location profile adds zero complexity to the selection of the algorithm. Clearly, more advance location profile can be used to further improve the achieved performance.

3.4.1.3 Location Profile and Location Estimation

In the case of location profile, if information is up-to-date and relevant to the current location of the device, traffic will be directed to the radio access in an optimal fashion. On the other hand, frequent updating of the information is costly. A more specific issue, is how mobile device monitors its locations. In outdoor scenarios, GPS provides a relatively precise estimation of the devices' locations but with an extra cost of the battery. In indoor scenarios, however, such precision does not exist and various methods for positioning of the device are explored in the literature. With the large deployments of the WiFi APs and small cells, WiFi fingerprinting is one of the interesting directions in this area [105]. In fact, many commercial buildings exploit WiFi fingerprinting for customised advertising to their visitors. Both location estimation and up-to-date location profiles as well as the cost associated with them, are open to further investigations.

3.4.2 Utility Definition

As mentioned earlier, two immediately available parameters at the device are battery status and RSS from various BS/APs. The aim of using utility function is to select the BS/AP with higher throughput while UE can have a lower power consumption at the same time. Therefore, a multi-criteria utility function, U_i as equation (3.13) has been defined, to allocate a utility value to the i^{th} BS/AP based on the considered criteria.

3.4. Device Controlled Network Selection

$$U_i = [u^{(1)}]^{w_1} + [u^{(2)}]^{w_2} \quad s.t. \quad w_1 + w_2 = 1 \quad (3.13)$$

where $u^{(1)}$ and $u^{(2)}$ are the utilities of RSS and battery, and w_k denotes weight values of these two utilities.

The same model as [106] has been used for defining the utility for RSS as equation (3.14),

$$u_x^{(1)} = \begin{cases} 0 & x < x_\alpha \\ \frac{(\frac{x-x_\alpha}{x_m-x_\alpha})^a}{1+(\frac{x-x_\alpha}{x_m-x_\alpha})^a} & x_\alpha \leq x \leq x_m \\ 1 - \frac{(\frac{x_\beta-x}{x_\beta-x_m})^b}{1+(\frac{x_\beta-x}{x_\beta-x_m})^b} & x_m < x \leq x_\beta \\ 1 & x > x_\beta \end{cases} \quad (3.14)$$

where

$$b = a(x_\beta - x_m)/(x_m - x_\alpha) \quad (3.15)$$

and

$$a \geq \max\left\{\frac{2(x_m - x_\alpha)}{x_\beta - x_m}, 2\right\} \quad (3.16)$$

where x_α and x_β represent upper and lower values of RSS. x_m is the middle value between x_α and x_β and a , b are the tuned steepness parameters. In the simulation, the value of a has been set to 2 [106].

The utility function of battery consumption is similar to [107], and according to equation (3.17).

$$u^{(2)} = e^{-\frac{(E_{ij} - E_i^{min})^2}{(E_i^{min})^2}} \quad (3.17)$$

where E_{ij} denotes battery consumption of the j^{th} device when it is connected with the i^{th} BS/AP, and E_i^{min} is the minimum value of battery consumption for connecting with the i^{th} BS/AP.

3.4. Device Controlled Network Selection

3.4.3 Reinforcement Learning

From the other aspect of optimizing values of utility functions, reinforcement learning mechanism has been used for network selection compared with that based on utility functions. The network selection algorithm mainly based utility function has been described in Algorithm 1.

Algorithm 1: NETWORK SELECTION BASED ON UTILITY FUNCTION

Data: BSs' and UEs' locations settings.

$U^{(1)}$: Utility values of RSS;

$U^{(2)}$: Utility values of battery consumption;

U_i : Total utility value combines RSS and battery consumption;

N_i : Index of different access points;

```
for  $t := 1$  to  $T$  do
    for  $j := 1$  to  $J$  do
        for  $i := 1$  to  $I$  do
             $[U^{(1)}]^{w_1}$  according to the equation (3.14);
             $[U^{(2)}]^{w_2}$  according to the equation (3.17);
             $U_i$  according to the equation (3.13);
        end
        The optimal selection for the UE is the alternative  $N_i$  with the
        highest  $U_i$  value;
    end
end
```

Objectives of using reinforcement learning in the simulation are maximising user-based Quality of Service (QoS) and saving their battery power. In addition, the benefit of combining reinforcement learning is that it can help UEs to consider historical records whenever performing handovers. UEs

3.5. Performance Investigations

can receive maximized cumulative reward values at different time steps of different kinds of networks [108]. Reinforcement learning is an incremental dynamic planning process, which can be used to determine the optimal strategy through step by step approach.

3.5 Performance Investigations

In this section, performance examination of the two different kinds of network selection approaches will be discussed. In the first part, the device centric MADM approach is studied which is focused on optimising throughput and delay to improve mobile users' quality level. In the second part, the device centric approach which considers history records by using reinforcement learning algorithm will be studied.

3.5.1 MADM Mechanism

In this section, simulation modelling of the proposed MADM traffic management algorithm will be described in detail to examine its performance. It modelled the network system with one LTE Macro cell, one pico and one femto, as well as one WiFi access point. It has been assumed that the coverage area of the three latter access points is included in the coverage area of macro cell. The quality level offered by these BSs/APs is a random value within the range listed in Table 3.1.

The required values by each application are also listed in Table 3.2, based on the values in [109, 110]. As explained earlier, the AHP algorithm has been used to weight the parameters for each application, as listed in Table 3.3. In this simulation, it has been assumed that different applications will focus on different performance criteria. Web browsing and video streaming are mainly considering their received throughput. Voice is considering delay as

3.5. Performance Investigations

Table 3.1: Quality offered by each radio access

Network	Bandwidth(Mbps)	Delay(ms)	Jitter(ms)
WiFi	11-54	100-200	15-30
LTE femto	10-100	50-100	15-25
LTE pico	10-100	50-100	15-25
LTE macro	10-100	50-100	15-25

its main performance metrics, while E-commercial has the same attention on these different performance criteria. Therefore, the criteria weight values of different applications are different.

Table 3.2: Boundary of quality level parameters depending on the application

Applications	Throughput(kbps)	Tx Delay(ms)	PacketSize
Voice	500-700	200-300	100KB
Web Browsing	300-600	100-300	10KB
Video Streaming	700-1000	300-600	2MB
E-commercial	600-800	50-100	10KB

There are 100 UEs, uniformly distributed in the full coverage area of the macro cell, i.e. a circle with diameter of 500 meters, in which the macro cell is located in the centre and the other three access points are randomly placed. Each of the 100 users are allocated to run a single application while an equal number of UEs run each type of applications (25 UEs run each of the applications). It has been assumed that the locations of UEs at 10 different time steps are randomly generated using uniform distribution in the range of (0, 500). The requirements of different applications are also randomly generated, based on uniform distribution within the range listed

3.5. Performance Investigations

Table 3.3: Weight values of different quality level parameters and for each application type

Application	Throughput	Delay	Jitter
Voice	0.0909	0.4545	0.4545
Web Browsing	0.6000	0.2000	0.2000
Video Streaming	0.6479	0.1222	0.2299
E-commercial	0.3333	0.3333	0.3333

in Table 3.2. Presented results are the average 10 runs of the simulation. However, the deviation among results of different runs were insignificant.

3.5.1.1 Simulation Cases

Three different cases have been considered for network selection by using MADM mechanism, as described below.

Case One: DC In this case, the selection of BSs/APs will be examined and it acts as a benchmark for the latter two cases. This case runs only based on the measurements at the device. In other word, this case is solely a device-centric scenario, hence it is DC case.

The value of throughput UEs received is been calculated based on Shannon Capacity as equations (3.18) and only free space path loss has been considered which affects value of Signal-to-Noise ratio (SNR) in the model.

$$\begin{aligned} Thr_{ij} &= B_i * \log_2(1 + SNR_{ij}), \\ Pl_{ij} &= 128.1 + 37.6 * \log_{10} d_{ij}. \end{aligned} \tag{3.18}$$

where j is number of UEs and i is number of alternatives in equations. d_{ij}

3.5. Performance Investigations

means the distance values between the j^{th} UE and the i^{th} BS/AP.

Except throughput UEs received, the end-to-end delay is also a significant factor of quality level metric. The end-to-end delay is generated by transmission delay and backhaul congestion together during the communication. The backhaul which connects BS/AP with the core network can be suffered by congestion because of high peak rates [111]. Therefore, the value of delay has been considered in this chapter is composed of two parts which are transmission delay and backhaul delay.

Case Two: DC/NA In this case, the device-centric/network-assisted (DC/NA) case is run which assumes mobile devices can communicate with the ANDSF server. Operator's policy and network view (e.g., congestion at different BSs/APs) are also taken into account while the network analytics mechanism can be moved to the device side. A set of random values has been generated for the past quality level offered by all BSs/APs within the range described in Table 3.1. It also generates random values in the range of $[10MB - 50MB]$ which represent the congestion level at each BS/AP. In this case, the network congestion has been translated to delay that can affect the end-to-end throughput respectively (a simple queuing model is assumed). When running the network selection approach, in this case, substitutes the A^+ and A^- of equations (3.7) and (3.8), with the best and worst possible alternative based on the history of the access points. The rest of the selection proceeds the same as case one.

Case Three:DC/NAA In this case, the network analytics has been used, in the form of location profiles, based on which best BS/AP can be selected. This case has been named as device-centric/network and analytic assisted (DC/NAA) case.

3.5. Performance Investigations

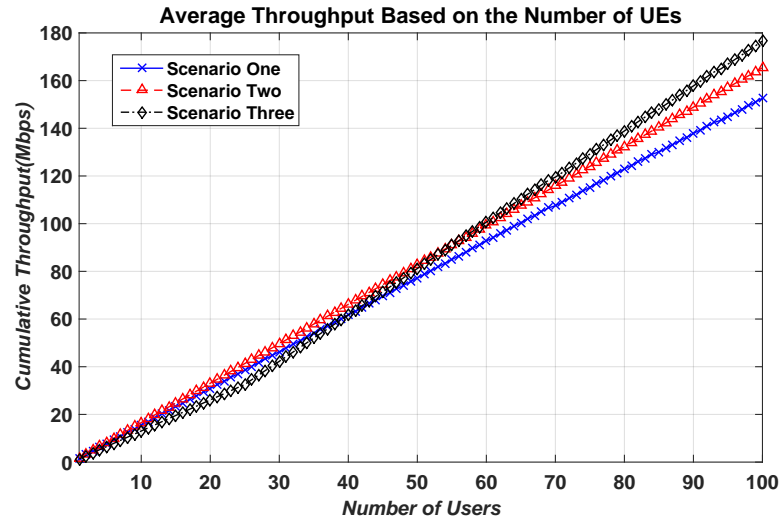


Figure 3.5: Total Average Throughput for UEs

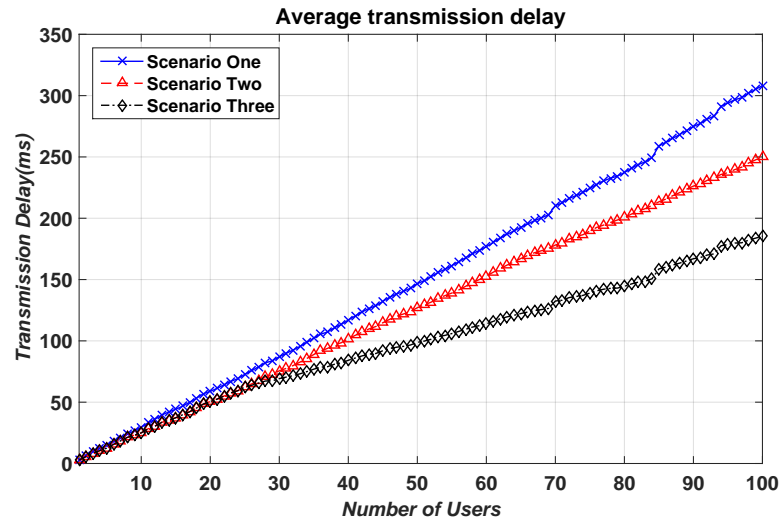


Figure 3.6: Total Average Delay for Mobile Devices

3.5. Performance Investigations

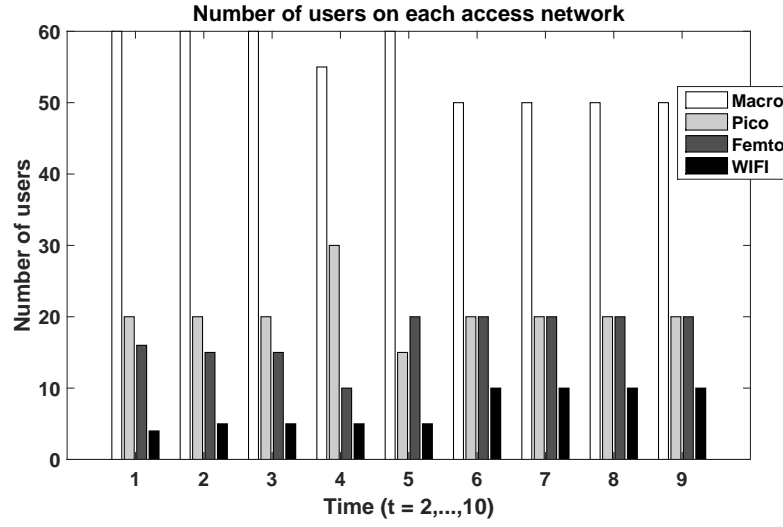


Figure 3.7: Balance of loads between different BSs/APs in the DC scenario (scenario one) at selected time stamps

3.5.1.2 Results Analysis

Fig 3.5 and Fig 3.6 show the aggregated values of average throughput and average delay per user in the three simulated cases. Observing from these two figures, moving from the first case that is DC only to the second and the third case that are DC/NA and DC/NAA, there is a small improvement in the throughput but a significant reduction in latency. This is mainly due to the fact that extra information on network congestion is integrated into the selection process in the two latter cases. Therefore, it can be argued that the enhancement on the end-to-end throughput (i.e. the goodput) is more significant in the DC/NA and DC/NAA cases.

It also looks into the balance of load between different BSs/APs in these three cases. Fig 3.7, Fig 3.8 and Fig 3.9 show the number of UEs connected to each of the BS/AP at different time stamps. These three figures show how the network information in case two can provide better balance of traffic distribution between BSs/APs, while relying only on the radio signal

3.5. Performance Investigations

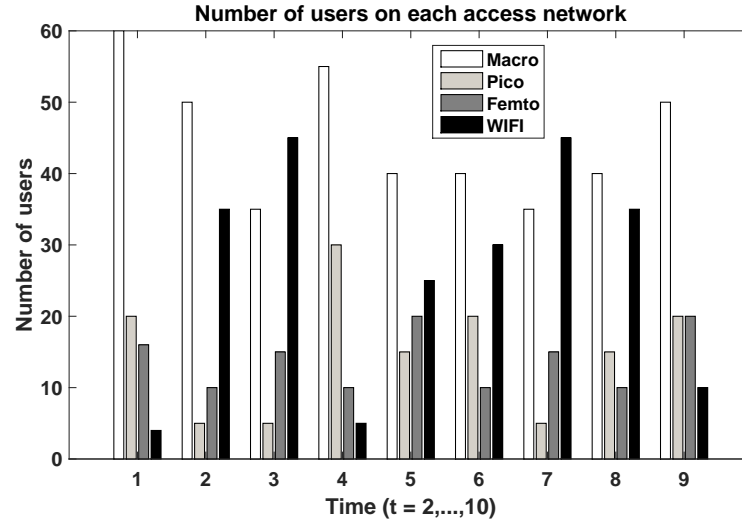


Figure 3.8: Balance of loads between different BSs/APs in the DC/NA scenario (scenario two) at selected time stamps

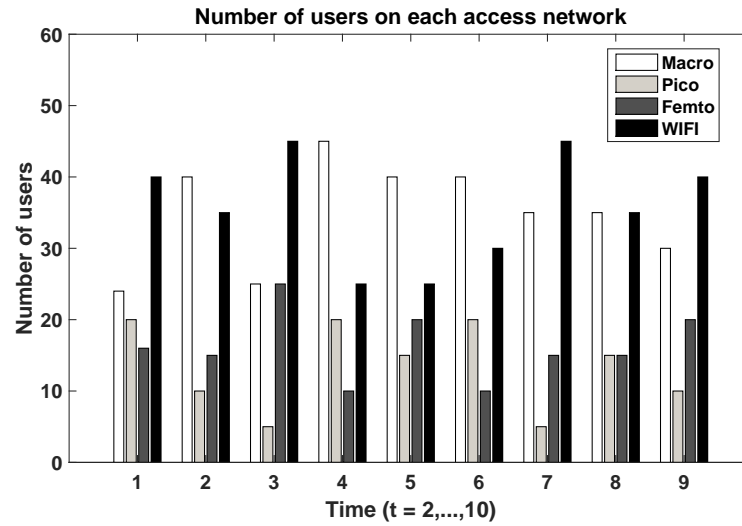


Figure 3.9: Balance of loads between different BSs/APs in the DC/NAA scenario (scenario three) at selected time stamps

3.5. Performance Investigations

measurements can result in potential congestion at some of the BSs/APs (the effect of which was seen in fig 3.6). The average values of standard deviation between the number of UEs connected to each BS/AP in each of the three cases, denoted by $\bar{\delta}$, is $\bar{\delta} = \{20.23, 17.91, 13.65\}$. Further observation from these three figures reveal that in the third case, since the decision on where to connect to is made based on long term statistics, the number of UEs who change their connected point throughout different time stamps of the simulation are less than the other two cases.

In summary, it can be seen that using the multi-objective optimisation (AHP and TOPSIS) at the device, enable mobile devices to make optimal network selection based on their QoS requirements.

3.5.2 Reinforcement Learning Mechanism

In this section, a simulation model of the proposed device-centric reinforcement learning based traffic management algorithm will be described in detail to examine its performance. The parameters of system model has been listed in Table 3.1 and Table 3.2.

In this simulation, network analytic mechanism with database has been designed the same as MADM mechanism. The database CNI which connects with ANDSF server stores information of history Q-values when UEs connect with the i^{th} BS/AP.

The backhaul congestion has been considered in this simulation as well, information of which has been stored in the database CNI. The received throughput of UEs will be affected by the quality of backhaul link. In order to model this, the Little's Law $\lambda_i = \frac{N_i}{T_i}$ has been used to generate values of backhaul capacity and congestion delay, where λ_i , N_i and T_i represent backhaul capacity, length of packets and waiting time in backhaul links.

3.5. Performance Investigations

3.5.2.1 Simulation Cases

Four different cases have been considered for network selection by using reinforcement learning mechanism, as described below.

Case One: *RSS-based selection* In this case, BSs/APs selection has been examined. It has been assumed that operator's policy and network analytic mechanism have been moved to UE side, and UEs can communicate with the ANDSF server directly. The history Q-values has been stored in database CNI, and then, Q-learning algorithm has been implemented with received signal strength (RSS) as its consideration.

Case Two: *battery-based selection* In this case, battery consumption of user equipment has been considered and Q-learning algorithm with network analytic mechanism has been implemented to select optimised RANs.

Case Three: *multi-criteria-based selection* In this case, Q-learning algorithm with network analytic mechanism has been implemented as well but consider RSS and battery consumption together. It aims to explain benefits of using multi-criteria metrics when doing access selection.

Case Four: *backhaul congestion-based selection* In this case, the same like previous scenarios but add backhaul congestion to the CNI. Because backhaul is one of the predominant factors which affects values of throughput UEs can receive. Such network analytic mechanism is crucially important for RAN selection, since e.g. a WiFi access point which offers relatively good connection quality at the radio level could be connected to a fully congested backhaul link. UEs received throughput will be affected by the quality of this radio link.

3.5.2.2 Results Analysis

Throughput UEs received from selected APs, energy consumption of battery and backhaul congestion level of UEs have been analysed in this section.

3.5. Performance Investigations

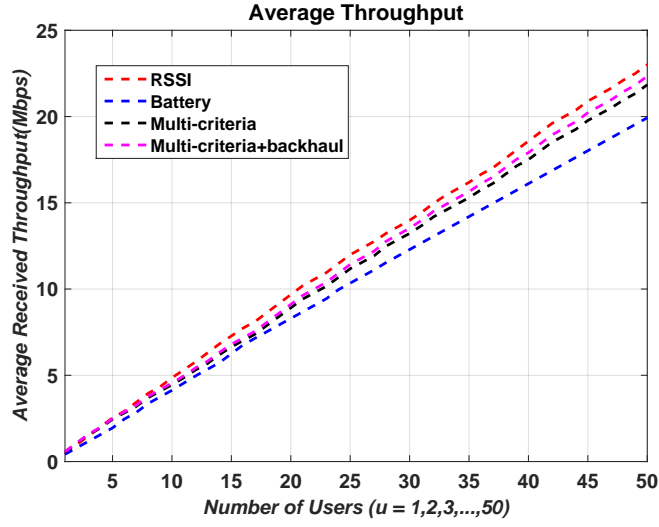


Figure 3.10: Average Cumulative Received Throughput

Fig 3.10 indicates cumulative average received throughput for UEs in four cases as we described above. From this figure, it can be explained that throughput values in case one are the highest, because UEs select APs only based on RSS. It means that the higher RSS, the better options for UEs to select. In case three, both RSS and battery consumption have been considered and weight values of these two criteria are 0.8 and 0.2, respectively. Therefore, UEs received throughput values are a bit lower compared with case one, but much higher than case two which only considers energy consumption of battery during the simulation. Throughput values in case four are a bit higher compared with case three, because history backhaul congestion has been considered which will help UEs to avoid the congestion link and receive more throughput.

Fig 3.11 indicates energy consumption of battery for UEs in four cases as we described above. From this figure, it can be explained that values of battery consumption are the lowest in case two. Because in case two mobile battery consumption has been considered as the main criteria. Case three and four generate the same value of battery consumption, that is because,

3.6. Conclusions

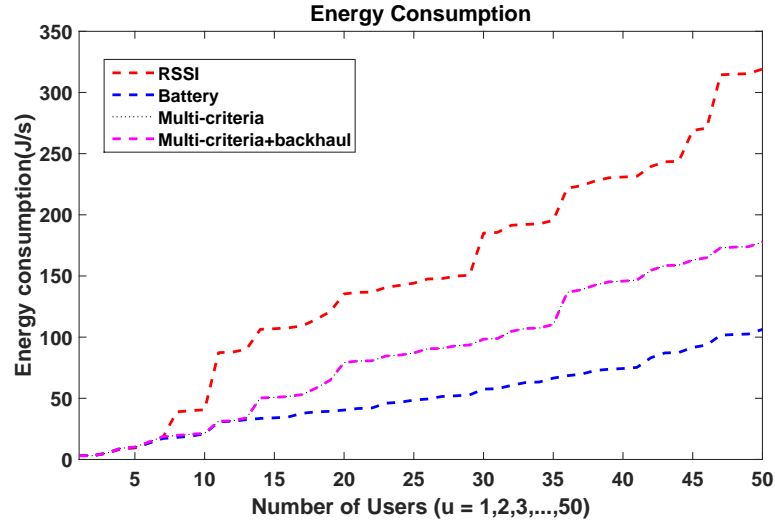


Figure 3.11: Battery-sensitive Energy Consumption

in case four, the considered backhaul congestion only affects values of UEs' received throughput without values of battery consumption.

Fig 3.12 indicates backhaul congestion level in four cases as described above. From this figure, it can be explained that case four generates the lest amount of backhaul congestion. It is because, in case four, UEs always consider history records of backhaul links, thus selected APs will provide more throughput and less congestion to maintain their communications compared with the other three cases.

3.6 Conclusions

In this chapter, fresh approaches for traffic management have been presented. The proposed mechanisms are device controlled and fully distributed at UE side. It benefits from the network analytic mechanism by communicating with the ANDSF server. Based on analysis results, it can be found that the proposed approaches can improve the received throughput, reduce the amount of battery consumption and the amount of average delay while im-

3.6. Conclusions

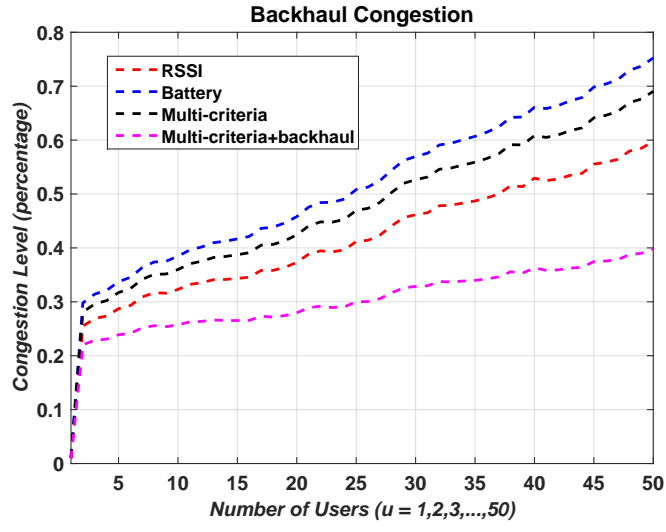


Figure 3.12: Utility Values in Different Cases

proving UEs' quality of satisfaction.

Chapter 4

Fairness Based Traffic Management in 5G

4.1 Introduction

The next generation of mobile networks, a.k.a. 5G, will be deployed with dense small cells of different technologies including LTE femtocells, LTE picocells, WiFi access points, and novel radios such as millimetre wave. Faced with an ever larger portfolio of applications to be served and with a corresponding number of requirements to be satisfied, it is commonly recognised that 5G networks need to consider various requirements of different application domains and industry sectors.

In order to address issues above, there has been a number of initiatives for the design of new mobile network architecture. One of the avenues for the 5G architecture design is the fully decoupled architecture. Decoupling of uplink and downlink has been well studied over the past few years [112] and its pros and cons are discussed in the community. Decoupling of the control and data plane is another well-investigated topic that is mostly studied within the context of Software-defined Networking (SDN) [113, 114]. Virtualisation

4.1. Introduction

and cloudification of mobile networking functionalities are another element of the 5G network that is enabled in such a decoupled architecture [115].

On the other hand and with the ever increasing data traffic in mobile networks, traffic management and maintaining Quality of Service (QoS) are more than ever challenging. According to Cisco Visual Networking Index, global mobile data traffic reached the 2.5 exabytes per month at the end of 2014 and this figure will surpass 24.3 exabytes by 2019 [116]. Hence, more efficient traffic managements are needed that can guarantee the QoS requirements from mobile users. In the SDN-based 5G network, it has been shown that centralised traffic management mechanism can provide guaranteed QoS and more efficient traffic management [117].

In this chapter, a traffic management mechanism based on *selfish* users and *fair* network will be defined, where the network side is a cloud-based central controller. The Jains fairness index has been used to quantify fairness [118], and simulated annealing algorithm is a kind of heuristic methods for solving the optimisation problem at the controller side. The Jains index has been well-used for quantifying the fairness in communication networks [111]. At the UE side, the problem of long-term QoS maximisation is formulated as a Q-learning problem based on the previous chapter.

The contributions of this chapter are twofold:

- A new QoS-based traffic management mechanism has been proposed which can maximise QoS of each UEs while maximising fairness index for the network system. In this chapter, the proposed mechanism will be named as QoS and Fairness maximisation (QFM). The Jains fairness will be used to define fairness among UEs and the fairness will be assumed to be maintained by the "central controller". Therefore, final decisions of network selection will be made by both UEs and the central controller together in order to maximise QoS utility of UEs as well as

4.2. Related Works

Jains fairness index between all traffic flows.

- The fairness levels of the system with QoS values constraints will be examined. It can be implemented by the system model which combines device-controlled mechanism and a cloud central controller together. In this chapter, a cloud central controller will be added on top of the whole system based on the previous chapter, and an optimisation approach will be devised to ensure traffic resources has been effectively managed based on the proposed QFM mechanism.

The remainder of this chapter is organised as follows. Section 4.2 briefly reviews the state of the art for cloud-based central controller and fairness in traffic management. After elaborating the system model and fairness approach in Section 4.3, the problem formulation and traffic management approach will be described in Section 4.4 and 4.5, respectively. Section 4.6 presents simulation study and performance observations. Finally, highlights of this work and road ahead are discussed in Section 4.7.

4.2 Related Works

In this section, the consideration of fairness in heterogeneous mobile networks will be reviewed. Fairness has been well-studied in the context of scheduling and wireless resource allocations, either on the wireless channel or over the end-to-end flow [119, 120]. Similarly, fairness has been studied in workload distribution in datacenters [121]. To quantify fairness, various different fairness measures have been proposed in literatures. The Jains fairness index [118], which was conceived to measure fairness in computer networks, is a very well used measurement of fairness in both wired and wireless networks [122]. A good fairness measure enables more mobile users to be accommodated with specific benefit requirements in the system. Therefore, the Jains

4.3. System Model

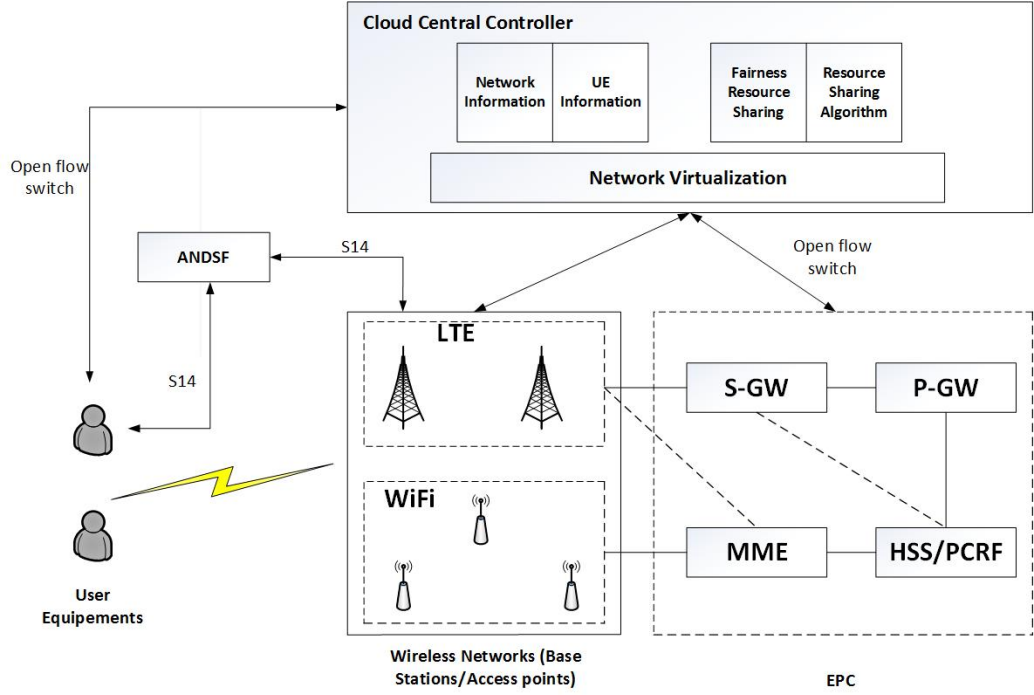


Figure 4.1: Cloud Central Controller based System Architecture

index has been used to measure fairness in this chapter as well.

4.3 System Model

Different elements of the proposed QoS and fairness maximisation (QFM) traffic management is detailed in this section which has been depicted in Figure 4.1. The three main layers in this model are the UE layer, the wireless network layer (i.e. radio access and mobile core, e.g. EPC), and the cloud layer (i.e. central controller). These three layers are elaborated in this section and the flow chart of communication between these layers has been described in Fig 4.2.

- **UE Layer:** In the system model, UEs can communicate with the ANDSF server [90] directly. It has been further assumed that the ANDSF includes network analytics server which can provide network analytic information to UEs through S14. The analytics considered

4.3. System Model

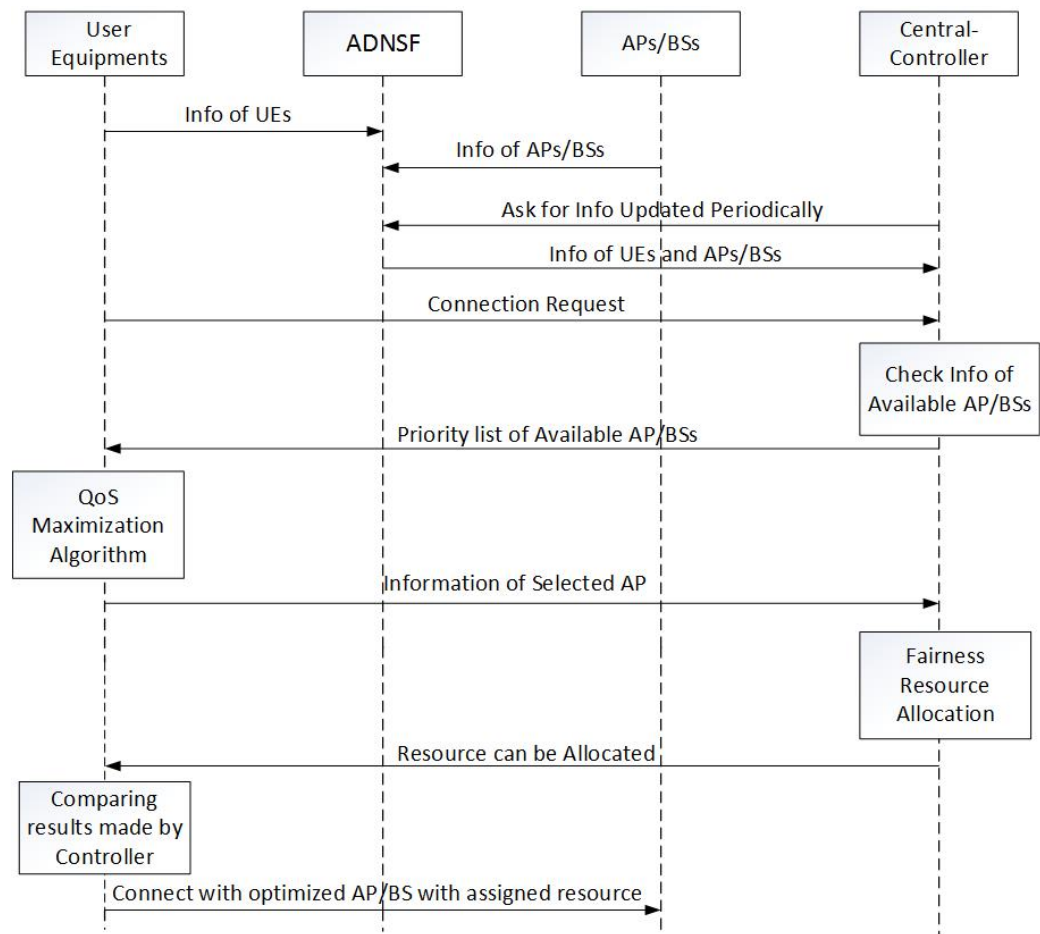


Figure 4.2: Communication Architecture

4.3. System Model

in this chapter is the performance of RANs in terms of QoS levels of UEs. This information has been stored in the database Candidate Networks Information (CNI) which connects with the ANDSF directly. The central controller communicates with UEs through open interfaces, i.e. the OpenSwitch. Therefore, decisions of fairness maximisation from the central controller have been made for UEs.

- **Wireless Network Layer:** The wireless network layer is made up of the radio access network and the mobile core network. It has been modelled that the radio access network includes four different Access Points (APs) which are LTE macro ($i = 1$), LTE pico ($i = 2$) and LTE femto ($i = 3$) as well as one WiFi access point ($i = 4$). The coverage area of these wireless access points is a circle with diameter R_i meter, where $R_i = \{500; 300; 50; 100\}$ (values from [99]). For simplicity, both of the WiFi AP and cellular base stations are referred to as wireless APs from this point on. The mobile core network consists of serving gateway (S-GW), packet data network gateway (P-GW), mobility management element (MME) and policy and charging rules function (PCRF) that have been used to implement connection, mobility and QoS management.
- **Cloud Central Controller:** The cloud-based central controller implements the following rules: (1) based on periodically updated information from UEs and APs, the central controller can check if any of the AP is available. If only one AP is available for a given UE, the controller will assign this AP to the UE; (2) if more than one AP is available for the UE, the controller will run fairness maximisation algorithm and provide the priority list of available APs to the UE.

4.4 Problem Formulation

In this section, details of optimisation problems will be presented which have been formulated for the proposed QMF mechanism. Two sets of problems which are the decision making by selfish users and fairness maximisation at the network will be described. Throughout the problem formulation, UE and user have been interchangeably used and it has been assumed that each user corresponds to a unique device.

4.4.1 Selfish Users

In the designed traffic management, UEs aim to maximise their quality level. The quality level will be defined as a utility function based on received throughput and the battery consumption, as equation (4.1),

$$\Phi = \frac{[\sum_i \sigma_{ij}]^{w_1} + [\sum_i E_{ij}]^{w_2}}{\sum_i \sigma_{ij} + \sum_i E_{ij}} \quad (4.1)$$

where σ_{ij} and E_{ij} represent values of received throughput, and consumed energy by the user $j = \{1, 2, \dots, U\}$ as a result of connecting with the AP $i = \{1, 2, \dots, N\}$. The first term in the numerator shows the total received throughput by user j , in case this user is connected with multiple APs. Similarly the second term is the total energy consumption by user j which has been described in Chapter 3. The w_1 and w_2 are weight values which represent the significance of these two different criteria in the utility function. We further assume that backhaul congestion affects UEs' received throughput. We model the backhaul link of each AP as a queue with exponentially distributed service time, $1/\mu_i$. Assuming $\overline{\sigma_{ij}}$ is the received throughput over the wireless channel (using Shannon equation), then the received throughput of user j through the i^{th} AP is $\sigma_{ij} = \min\{\overline{\sigma_{ij}}, \mu_i\}$.

4.4. Problem Formulation

To this end, the following optimisation problem will be solved by each UE, i.e. UE_j as equation (4.2).

$$(P1): \text{maximize } \Phi_j \quad (4.2)$$

$$\text{subject to, } \sum_i E_{ij} \leq P_j, i = \{1, 2, \dots, N\} \quad (4.2a)$$

where P_j denotes the remaining battery at the j^{th} UE.

4.4.2 Fair Network

The cloud-based controller aims to maximise fairness among all connected UEs. As explained earlier, the Jains fairness index can be used to quantify the achieved fairness among UEs as a result of the proposed traffic management mechanism. Jains fairness index can be explained in equation (4.3),

$$J(X) = \frac{(\sum_{j=1}^U x_j)^2}{U * \sum_{j=1}^U x_j^2} \quad (4.3)$$

where $x_j = \sum_i \sigma_{ij}$, that is the total received throughput by the UE_j .

Therefore, the optimisation problem at the cloud-based controller can be formulated as equation (4.4),

$$(P2): \text{maximize } J(X) \quad (4.4)$$

$$\text{subject to, } x_j \geq R_j^{min}, \forall j \in \{1, 2, \dots, U\} \quad (4.4a)$$

where R_j^{min} shows the QoS requirement (in this case minimum required throughput) of the user j depending on its application.

4.5 Traffic Management Solutions

In this section, different solutions for the proposed traffic management optimisation problems will be presented in **Section 4.4**. First, Q-learning will be used to solve the selfish users' optimisation problem (P1) and then simulated annealing will be used as a meta-heuristic to solve the fairness maximisation problem (P2).

4.5.1 Solutions to (P1) Using Q-learning

In order to solve (P1) for QoS optimisation, Q-learning has been used. The main reason for using a learning based approach is the possibility of including historical data so as to make decisions that is optimal choices for UEs in a longer period of time, and to potentially reduce the number of handovers.

Q-learning is an incremental dynamic planning process, which can be used to determine the optimal strategy through step by step approach. Therefore, time-varying states, actions and reward functions for the process of selecting the optimal AP should be defined. At each time t , $s(t)$ describes the state of a given AP, which will alter to $s(t+1)$ by executing action $a(t)$. The Q-value of this transition is defined as the expected value in equation (4.5).

$$Q^t(s, a) = E\{R_t | s = s(t), a = a(t)\}. \quad (4.5)$$

The state $s(t)$, the action $a(t)$ and the reward value of $R(t)$ have been defined as below

- $s_i(t)$: State of AP_i at time t is denoted by $s_i(t) \in S$ and represents receiving service through AP_i .
- $a_i(t)$: Actions $a_i(t) \in A$ represents changing from one AP to another.

4.5. Traffic Management Solutions

- $R_i(t)$: We define the *Reward Function* based on the value of Ω in equation (4.1). Equations (4.6) and equation (4.7) describe the immediate reward function, $r_i(t)$, and the weighted aggregated reward function over time, $R_i(t)$,

$$r_i(t) = (\sum_i \Phi_j(t) - \sum_i \Phi_j(t-1)), \quad (4.6)$$

$$R_i(t) = \sum_{k=1}^{10} \gamma^k r_i(t-k), \quad (4.7)$$

where k demonstrates number of historical records that are taken into account and γ is the discount factor. In other words, γ represents significance of previously recorded reward values of $R_i(t)$. In the simulation study of this chapter, we set $\gamma = 0.995$ and the time stamps, $t = 10$, in equation (4.7) is similar to the described algorithm in [123].

Based on parameters described above, Q-values can be calculated by considering historical records as equation (4.8),

$$Q^t(s, a) = Q^t(s, a) + \alpha[R(t) + \gamma Q^t(s, a) - Q^{t-1}(s, a)] \quad (4.8)$$

where $Q^t(s, a)$ is the current value of Q for a given AP at time t , and $Q^{t-1}(s, a)$ is historic values that have been stored in the CNI and retrieved by the UE. Parameter α represents the learning rate, that is a value in the range of $(0, 1)$, if $\alpha = 0$, the Q value is never updated. Summary of the Q-Learning algorithm for solving (P1) is described in Algorithm 2.

4.5.2 Solutions to (P2) Using Simulated Annealing

Simulated Annealing (SA) is a well-used heuristic for solving combinatorial problems. At each step of the SA algorithm, current solution will be replaced

4.5. Traffic Management Solutions

Algorithm 2: QoS MAXIMISATION BASED ON Q-LEARNING

$\sum_i \Phi_j$: total utility value of the i^{th} network

$r_i(t)$: immediate reward value at time t of the AP i

$R_i(t)$: cumulative reward value at time t of the AP i

$Q_i(t)$: cumulative Q value at time t of the AP i

γ : discount rate which reflects the values of importance of rewards

α : learning rate and equals to 0.01

```
for  $t := 1$  to  $T$  do
  for  $i := 1$  to  $I$  do
    for  $j := 1$  to  $J$  do
      | Calculate  $\sum_i \Phi_j$  for each network;
    end
    Calculate immediate reward value according to the function
      
$$r_i(t) = \sum_i \Phi_j(t) - \sum_i \Phi_j(t - 1);$$

    Calculate cumulative reward function  $R_i(t)$  according to the
      function
      
$$R_i(t) = r_i(t) + \gamma r_i(t - 1) + \gamma^2 r_i(t - 2) + \dots = \sum_{k=1}^{\infty} \gamma^k r_i(t - k);$$

    Then, the Q-values we are planning to get are received by
      function  $Q_i(t) \leftarrow Q_i(t) + \alpha[R_i(t) + \gamma Q_i(t) - Q_i(t - 1)]$ 
    end
    UEs will select the network  $i$  with biggest  $Q_i(t)$  value at different
      time steps.
  end
end
```

4.6. Performance Investigations

with a new solution by a certain probability. That probability depends on both difference between the current solution and the randomly generated neighbour solution, and also the temperature value F of the system [107].

In this section, a solution for (P2) based on simulated annealing algorithm will be described. It runs at the central controller and has been detailed in Algorithm 3. The algorithm aims to maximise achieved fairness values among UEs where the Jains fairness index has been implemented as explained in 4.4.2 to quantify fairness. Solving (P2), using the SA algorithm to consider maximising $J(X)$. If the value of $J(X)$ of the neighbour AP is higher than that of the current one, the algorithm triggers a move to the neighbour AP. Otherwise, the algorithm will choose an AP between the current AP and the neighbour AP according to a generated probability value. The random selection will allow the solution to converge to a global optimal point. The generated probability value of replacing the current AP to the neighbour AP is based on $p = \frac{\Delta J(X)}{F}$, where $\Delta J(X)$ indicates the difference of $J(X)$ value between the current AP and the neighbour one.

4.6 Performance Investigations

In this section, the simulation settings, discussions, and analysis of simulation results will be explained.

4.6.1 Simulation Parameters

As mentioned earlier, the simulation system has been modelled as an integrated wireless network. The wireless network channel is modelled with path loss (see Table 4.1), and hence the RSS can be explained as $RSS_i = Pt_i - PL(d_i)$, where Pt_i denotes the transmit power of the i^{th} AP, d_i is the distance between UE and the i^{th} AP and $PL(d_i)$ is the associated path loss

4.6. Performance Investigations

Algorithm 3: TRAFFIC MANAGEMENT AT THE CLOUD-BASED CONTROLLER

Data:

$J(X)$: Jain's index value;

F_0 : initial temperature;

κ : temperature decrease rate, where $\beta \in (0, 1)$;

$\bar{\kappa}$: change limitation value;

for $T := T_0$ **to** T_n **do**

 Compute current value of Jain's index $J(X)$;

 select neighbour AP_i randomly, and calculate neighbour Jain's index value $J(X)'$ while make sure the value of x_j meets the minimum requirements;

if $J(X)' - J(X) > 0$ **then**

 the i_{th} neighbour AP becomes current result;

 received throughput value of neighbour σ'_{ij} replaces current throughput value σ_{ij} as well;

else

 generate accept probability $p = \frac{-\Delta J(X)}{F}$;

 then, the i_{th} neighbour AP replaces current AP based on the probability p ;

end

end

$F_n = \kappa * F_{n-1}$, if the fairness value $J(X)$ is not changed more than $\bar{\kappa}$, then stop.

end

4.6. Performance Investigations

function [101]. Detailed simulation parameters are described in Table 4.1.

Table 4.1: Main Simulation Parameters

Parameter		Value
Peak Data Rate	LTE	100Mbps
	Wi-Fi	11Mbps
Tx Power	LTE macro	46dBm
	LTE pico	23dBm
	LTE femto	13dBm
	Wi-Fi	20dBm
Noise Spectral Density		-174dBm/Hz
Path Loss Model		$128.1 + 37.6 * \log(d)$
App Throughput	Video	500-700 kbps
	Interactive	300-600 kbps
	P2P	700-1000 kbps
	E-service	600-800 kbps
Cell Coverage	LTE macrocell	500m
	LTE picocell	300m
	LTE femtocell	50m
	WiFi	100m

4.6.2 Simulation Cases

- *Case One (QoS-based traffic mechanism):* In this case, QoS-based RAN selection without considering history records has been examined. Weight values of different criteria have been set as 0.8 for received throughput and 0.2 for consumed energy by UEs, respectively. It has been assumed that decisions have been made by UEs and UEs can

4.6. Performance Investigations

communicate with the ANDSF directly in order to receive information of candidate APs as described in Section 4.3. Therefore, the UE selects the AP which can offer a highest value of QoS based on equation (4.1).

- *Case Two (learning-based traffic mechanism)*: In this case, reinforcement learning based RAN selection has been examined. Compared with case one, Q-learning algorithm runs at UEs side by considering history Q-values of each available AP. This is based on solving P1 as explained in **Section 4.4.1**. The main aim of using history values is to reduce the potential number of handovers by selecting the AP that has high performance over a period of time (and not only instantaneously).
- *Case Three (QFM-based traffic mechanism)*: In this case, the proposed QoS and Fairness maximisation (QFM) based RAN selection has been examined. In case three, it solves (P1) by using Q-learning at the UEs side and solves (P2) by using simulated annealing at the network controller side. If there exists a conflict between results from the UE side and the network controller side, the received throughput values of UE_j should be checked simultaneously. If received throughput values that generated from the network controller side are in the field of application throughput described in Table 4.1, then the process of selecting APs is based on the results generated from the network controller side. Otherwise, the selected APs are based on the results generated from the UEs side.

4.6.3 Result Analysis

The considered Key Performance Indicators (KPI) are: UEs throughput, UEs battery consumption, number of handovers and Jains fairness index.

Figure 4.3 indicates aggregated average received throughput for UEs in

4.6. Performance Investigations

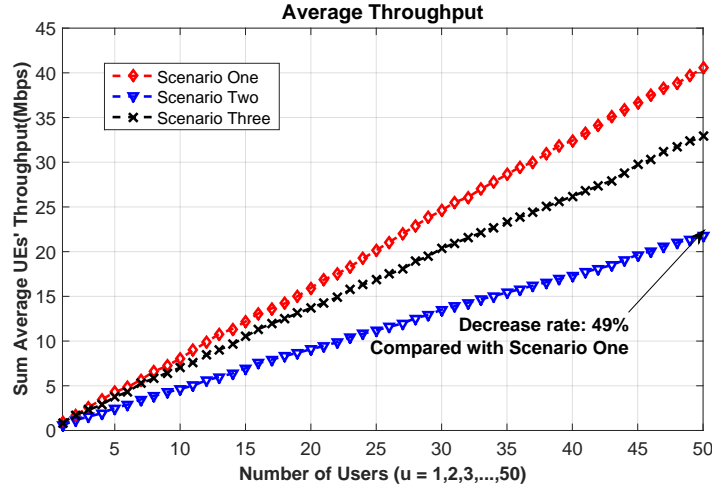


Figure 4.3: Sum throughput Vs. number of active users

three different cases as we described above. Observing from this figure, it can be shown that throughput values in case one are the highest. That is because, the aim of QoS-based traffic mechanism in case one is maximising QoS values based on equation (4.1). Maximising QoS values means maximising the value of UEs received throughput. Since backhaul congestion are also considered in case one, the UE is able to connect with an AP that provides higher throughput (and not only higher data rate over the wireless channel). Therefore, the value of total average throughput UEs can receive in case one is higher than that in case two and three.

In case two, the total average received throughput of UEs is approximate 50% lower than that of case one. That is because, the main aim of learning-based traffic mechanism is to reduce the potential number of handovers without focusing on improving UEs received throughput. Therefore, the value of throughput decreased dramatically and from this figure, it can be indicated that not all UEs can receive their required throughput values.

In case three, it can be explicitly shown that the total average received throughput of UEs is lower than that in case one, but higher than that

4.6. Performance Investigations

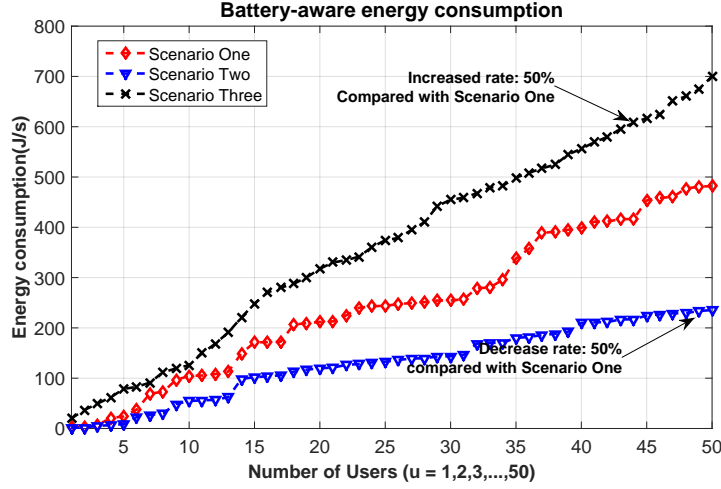


Figure 4.4: UEs' average battery consumptions

in case two. That is because, the aim of QFM-based traffic mechanism is improving fairness allocation for all UEs which is restricted by achieving minimum required throughput of UEs. Therefore, the value of total received throughput should be increased compared with case two, but still decreased compared with that in case one which is mainly focusing on maximising throughput values.

Figure 4.4 shows battery consumption of the UEs in three different cases. It can be shown that values of battery consumption in case two are the lowest, and those values in case three are the highest. Because in case two, Q-learning algorithm has been implemented at the UEs side. Cumulative reward values can help UEs to learn from history experience of candidate networks and can help them to perform the best action at each time steps. The aim of using Q-learning is to reduce the number of handovers for all UEs to help them to maintain their ongoing communications for a longer period of time. Therefore, the value of battery consumption should be reduced in case two and it is lower compared with that in the other two cases.

The reason why battery consumption in case three is higher than that

4.6. Performance Investigations

Table 4.2: Total Number of Handovers in Each Scenarios

	Scenario One	Scenario Two	Scenario Three
Total No. of Handover	209	172	257

in case one is the more handover occurred. As described in the equation (3.2) of **Section 3.3.2**, device energy consumption is based on three different variables which are d, τ and T . In these three cases, values of τ are the same. The more number of handovers, the more time T wasted. From Table 4.2, the total numbers of handovers are 209, 171 and 257 in these three cases, respectively. Therefore, in Fig 4.4, decreased and increased values of energy consumption are proportional to case one. The more number of handovers generates the more battery consumption.

Number of handover over the course of simulation are demonstrated in Fig 4.5, Fig 4.6, and Fig 4.7. It can be seen that average number of handovers in case two (Fig 4.6) is lower than those in case one (Fig 4.5). Reduced number of handovers for UEs can decrease values of battery consumption and can confirm the result presented in Fig 4.4. Higher number of handovers can be observed in Fig 4.7 that correspond to the higher battery consumption of case three in Figure 4.4.

Furthermore, Fig 4.8 shows the QoS utility value (as equation (4.1)). It can be shown that the QoS utility value is lowest in case three and highest in case one. That is because, the aim of case one is enabling UEs to connect with optimal APs which can provide maximum value of QoS. Based on equation (4.1), though values of energy consumption are higher in case one, the weight value of it is much smaller compared with that of throughput which is the main affect factor for QoS level. Therefore, QoS level in case two is lower than that in case one, even though the energy consumption has been reduced

4.6. Performance Investigations

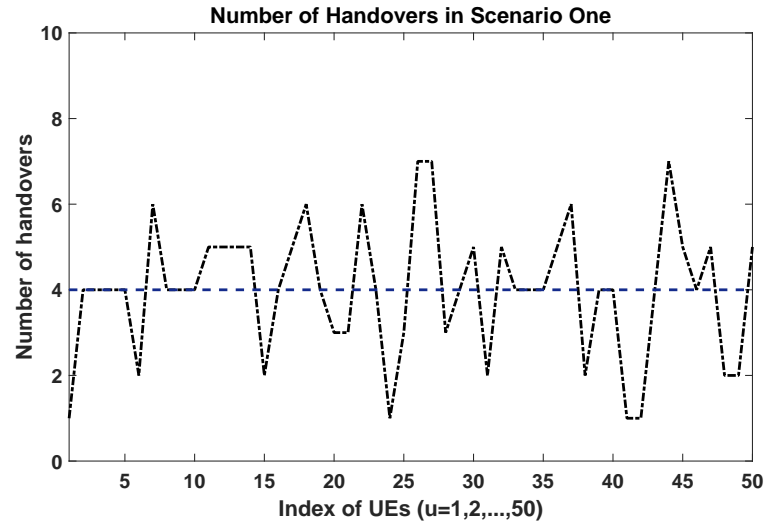


Figure 4.5: Number of Handovers in Scenario One (QoS-based)

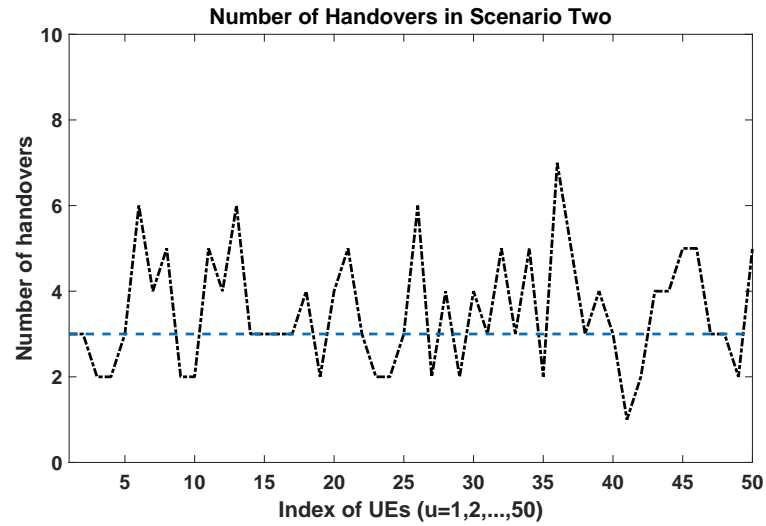


Figure 4.6: Number of Handovers in Scenario Two (Learning-based)

4.6. Performance Investigations

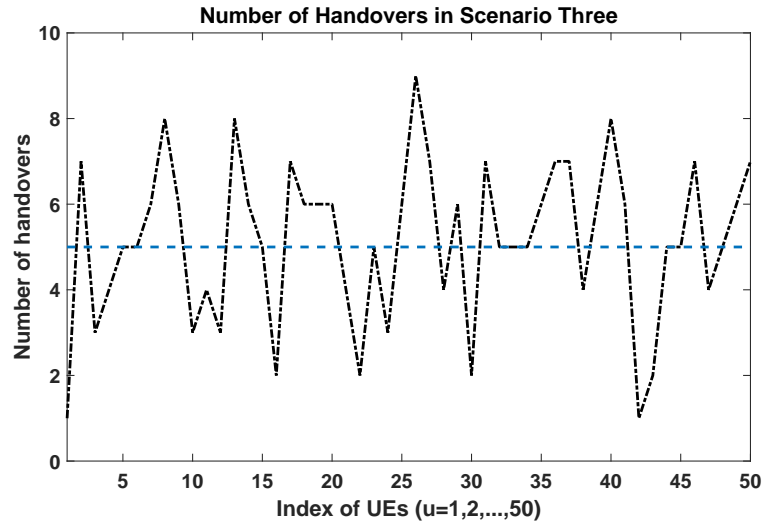


Figure 4.7: Number of Handovers in Scenario Three (QFM-based)

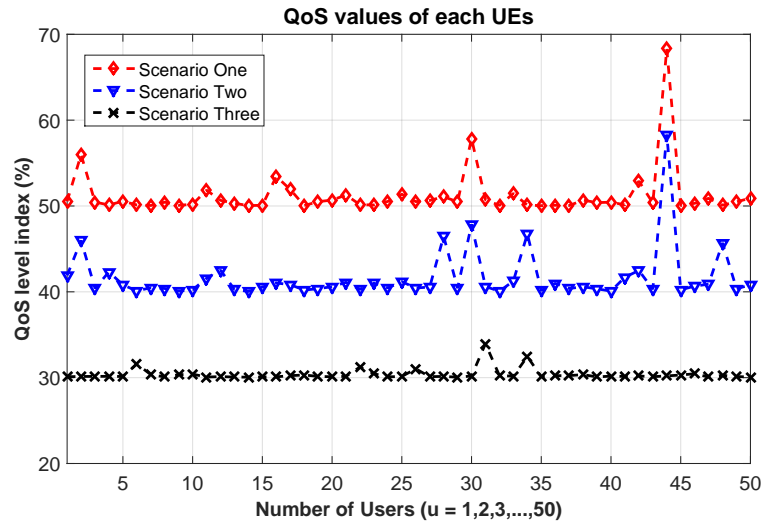


Figure 4.8: QoS Utility based on Equation 4.1

4.6. Performance Investigations

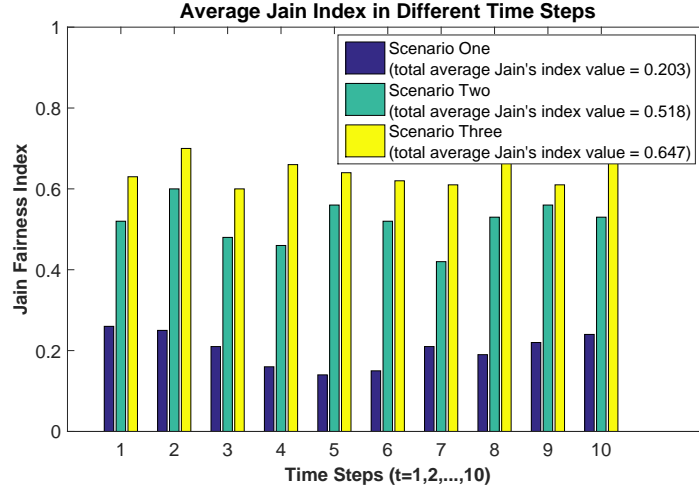


Figure 4.9: Average Jain Index Value vs. Simulation Time Steps

to a large extent. In case three, our proposed QFM mechanism implements fairness traffic mechanism which can reduce values of QoS for UEs at the same time. After UEs select APs by considering their selfish requirements, our cloud central controller reassigns traffic resource to UEs with respect to fairness.

Finally, Jain's fairness index is plotted in Fig 4.9. As expected, case three has the highest fairness index and case one has the lowest fairness index. Based on fairness equation (4.3), the value of $J(X)$ is in the field of $(0, 1)$ and the higher the better. Higher $J(X)$ value will be generated when values of $(\sum_{j=1}^U x_j)^2$ and $(U * \sum_{j=1}^U x_j^2)$ are quite similar. In case three, fairness problem (P2) has been solved by our proposed QFM mechanism. Network resource has been allocated efficiently while minimum requirements of UEs have been satisfied as well. In case one, each UEs purchases higher throughput and connects with the AP which can provide the highest resource at the same time. Therefore, values of throughput will be quite different between UEs and it will generate the lowest fairness index in case one.

4.7 Conclusions

In this chapter, a novel approach to traffic management in heterogeneous mobile networks has been presented, which is QoS and Fairness Maximisation (QFM) mechanism. With the rapidly increasing number of mobile devices, their throughput demand and longer battery lifetime requirements, maximising their QoS levels will be the significant part in the next generation networks. Meanwhile, how to allocate traffic resources in a fairness way is another important issue for us to consider. The proposed QFM mechanism is composed of two parts which are fully distributed QoS maximisation mechanism at UE side, and centralised fairness traffic management mechanism at controller side. These two parts are implemented by UEs and cloud central controller separately with the whole view of the system. Based on analysis results, it can be found that the proposed fairness problem have been solved. Resources of the network have been fairly allocated and the fairness index has been maximised as well.

Chapter 5

Network Slicing based Traffic Management in 5G

5.1 Introduction

With the fast growth of wireless network technologies (e.g. 5G) and ever-increasing demand for services with high Quality of Service (QoS) demand [124], the management of network resources becomes an always more challenging task that needs to be properly designed in order to improve network performance. In this scenario, network slicing [125] is gaining an always increasing importance as an effective way to introduce flexibility in the management of network resources. A slice is a collection of network resources, selected in order to satisfy the requirements (e.g., in terms of QoS) of the service(s) to be provided by the slice [126]. The intention of slicing is to introduce flexibility and higher utilisation of network resources by providing only the network resources necessary to fulfil the requirements of the slices enabled in the system.

An enabling aspect of network slicing is the virtualisation of network resources, which allows operators to share the same physical resources in

5.1. Introduction

a flexible, dynamic manner in order to exploit the available resources in a more efficient way [127]. Virtualisation of network resources is currently investigated in literature especially by focusing on the virtualisation of network functionalities [127, 128, 129, 130]. Due to the diverse QoS requirements and the limitation of network resources, efficiently allocate network resources among service slices and user equipments (UEs) is a significant issue [131]. In this field, further study is needed on the virtualisation of radio resources in order to perform the admission control and the resource allocation for network slices. Indeed, an important aspect to be considered is the way radio resources are allocated to different slices in order to meet requirements of such slices. The task relevant to radio resource allocation becomes more challenging with network slicing, as it introduces a two-tier priority in the system. The first tier, inter-slice priority, refers to the priority of different slices, as each slice has its own priority defined according to the agreements between the slice owner and the network provider. The second tier, intra-slice priority, refers to the priority among the users of the same slice.

When looking at the solutions exploited over current 4G systems to manage radio resources, it clearly emerges that 4G networks are able to maximise the QoS of the served users but are not able to perform the resource allocation in slicing environments [132]. Therefore, UEs will receive resources from the same traffic with the same priority value. This limitation is due to the fact that resource allocation in 4G systems is performed by associating a priority to the service requested by the UEs. This approach thus fails when considering that in 5G systems different UEs may belong to different slices with different priorities, and thus such UEs should be managed by considering the priority of the slice they belong to plus the priority of the service they require. Therefore, UEs in the 5G system will get higher QoS compared with

5.1. Introduction

4G systems because the same traffic requirement can be satisfied by different slices with different priority values.

In this chapter, a novel heuristic-based admission control mechanism has been proposed. As shown in Fig. 5.1, the proposed admission control mechanism exploits a two-tier priority level. The proposal is based on the idea that network slices communicate to an admission control entity the desired QoS level. The admission control mechanism, based on the priority of the slice, decides about serving the slice. Finally, according to the inter- and intra-slice priority, the virtual network allocates the physical radio resources to the UEs of admitted slices. According to the decision of admission control, the resource allocation task is performed with the aim to maximise the quality level of the users within each slice, by considering the inter-slice priority. In this chapter, the quality level is measured by considering the effective throughput experienced by the users, normalised according to their maximum requested data rate. With this aim, the resources allocated to a slice with low priority could be reduced, if necessary, down to the minimum amount able to meet the basic QoS requirements in order to admit new slice(s) with higher priority. So doing, the proposal dynamically changes the amount of network resources allocated to network slices according to the traffic load without affecting the quality level of the users and while improving the network utilisation. To summarise, the main contributions of this chapter can be stated as follows:

- A novel heuristic based admission control mechanism with two-tier priority level have been proposed in our virtualised 5G system model. The proposed admission control mechanism dynamically set the resources allocated to enabled slices according to the current traffic load.
- Inter-slice and intra-slice priority order has been taken into account for designing the quality level maximisation problem of resource allocation

5.2. Related Works

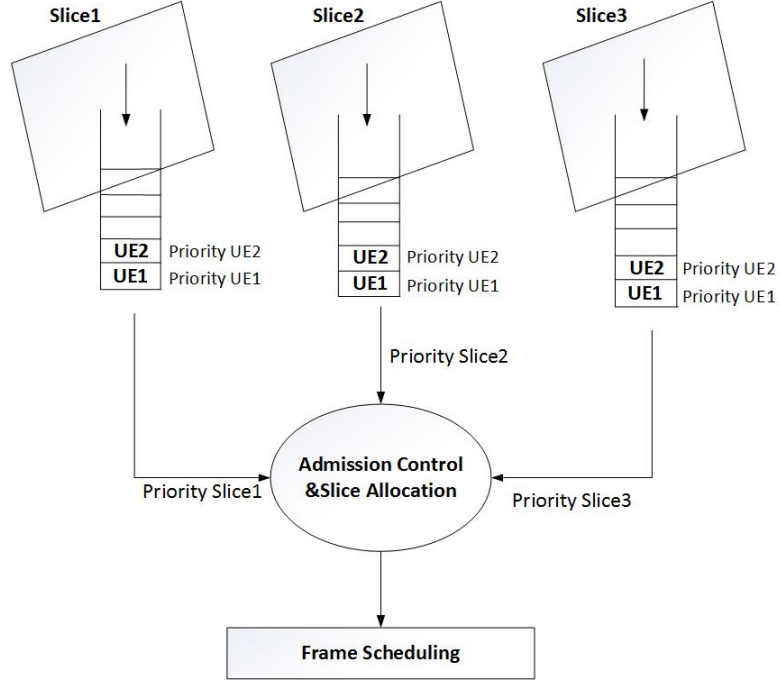


Figure 5.1: Our reference scenario with inter-slice and intra-slice priority.

task. Considering priority orders for the quality level function can improve the satisfactory level of UEs and network utilisation.

The remainder of this chapter is organized as follows. Section 5.2 reviews the state of art on virtual resource allocation in different kinds of network technologies. After elaborating the system model in Section 5.3, the proposed admission control mechanism will be described in Section 5.4. Section 5.5 presents simulation study and performance observations. Finally, summary of this chapter is given in Section 5.6.

5.2 Related Works

In literature, several solutions for efficiently supporting virtualisation of network resources have been designed to improve the quality level of UEs and network resource utilisation [129].

An efficient wireless network virtualisation for Long Term Evolution (LTE) systems has been proposed in [133], which proposes a slicing scheme to effi-

5.2. Related Works

ciently allocate physical resource blocks to different service providers (SPs) in order to maximize the utilisation of resources. The scheme is dynamic and flexible for addressing arbitrary fairness requirements of different SPs. Similarly, [134] proposed a framework for wireless resource virtualisation in LTE system to allow sharing of radio resources between mobile network operators. An iterative algorithm has been proposed to solve the Binary Integer Programming (BIP) with less computational overhead. Nevertheless, above considered schemes do not consider the priority among different slices as well as the priority among the users within the same slice.

For the limitation of network resources, the admission control mechanism can be implemented to improve communication reliability and network utilisation. In [70], a joint resource provisioning and admission control mechanism have been proposed aiming to maximise the total rate of virtualised networks based on their channel state information. An iterative slice provisioning algorithm has been proposed to adjust minimum slice requirements based on channel state information but without considering global resource utilisation of the network as well as inter- and intra-slice priority.

In [135], a mechanism for allocating downlink network resources has been proposed. The mechanism decides to accept a novel service only if the provisioning of this new service does not affect the throughput of the services in the cell. As a consequence, this work does not take into consideration the dynamic modification of the quality level experienced by mobile users in order to increase network capacity and resource utilisation.

Centralised joint power and admission control mechanism for prioritised multi-tier cellular networks has been proposed in [78]. The mechanism has been developed to admit users with higher priority level in order to maximise the number of users. In this case, the priority is only considered at the user level and, thus, this work fails in guaranteeing differentiation in case users

5.3. System Model

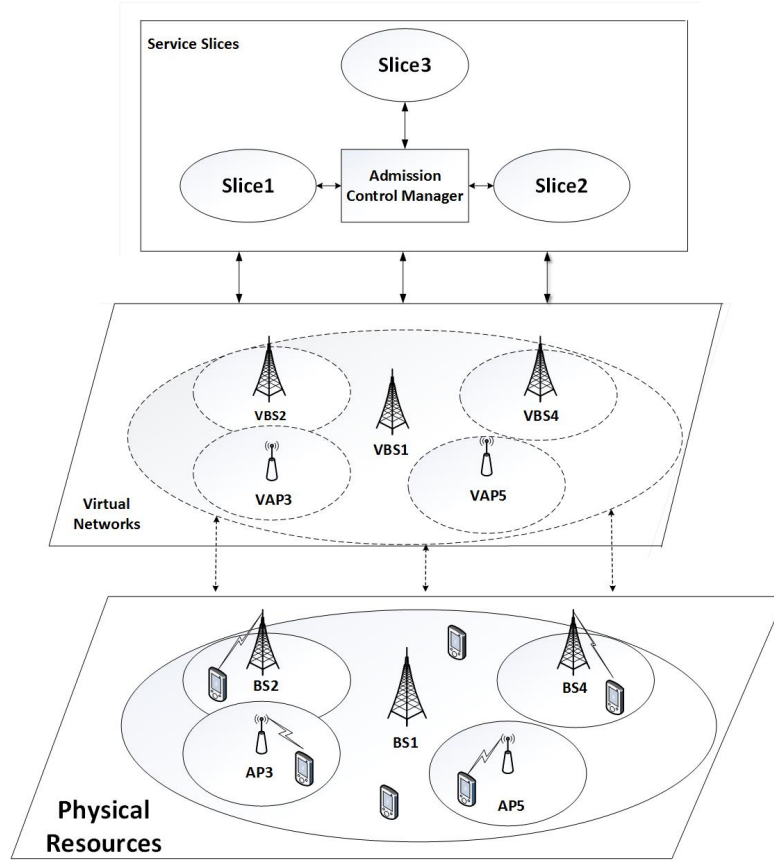


Figure 5.2: System Architecture

belong to slices with different priorities.

5.3 System Model

As depicted in Fig. 5.2, the model consists of four main elements: the service slice layer, the virtual network layer, the physical resources, and the admission control manager. The first three elements will be explained in the remainder of this Section, while the admission control manager will be treated in Sec. 5.4.

5.3. System Model

5.3.1 Service Slices

The service slices present different services (e.g., car management, TV streaming and web browsing) which require resources to be served. We indicate with $\mathbb{S} = \{1, 2, 3, \dots, S\}$ the set of slices in the virtual network. Each slice s has a set of UEs, such a set is denoted by $\mathbb{U}_s = \{1, 2, \dots, U_s\}$. Each slice s performs a request to the admission control in terms of QoS constraints. In this chapter, it models such a request with R_s^{min} and R_s^{max} , which denote the minimum and maximum data rates associated with the slice s , respectively.

Each slice s is characterized by a priority, ρ_s , where such priorities are defined with the constraint that $\sum_{s \in \mathbb{S}} \rho_s = 1$. Similarly, each user u belonging to the slice s , i.e., u_s , is characterized by a priority μ_{u_s} , where $\sum_{u_s \in \mathbb{U}_s} \mu_{u_s} = 1$.

5.3.2 Virtual Network

The virtual network layer provides an abstraction of the physical network resources. According to the decisions of the admission control, the virtual network slices the resources of network to accommodate different slices. The virtual network receives the requests of different slices in terms of UEs to be served for each slice, and performs the subsequent allocation of physical resources according to the inter- and intra-slice priority while taking into account the quality level of UEs.

With this aim, we can define:

$$\phi_{u_s} = \left(\frac{r_{u_s}}{R_s^{max}} \right) \quad (5.1)$$

as the quality level of UE u in the slice s ; r_{u_s} is the data rate of the UE u in the slice s . The overall quality level of users belonging to slice s can be

5.3. System Model

computed as:

$$\phi_s = \sum_{u_s \in \mathbb{U}_s} (\phi_{u_s})^{\mu_{u_s}} \quad (5.2)$$

Finally, we can define:

$$\Phi = \sum_{s \in \mathbb{S}} (\phi_s)^{\rho_s} \quad (5.3)$$

as the overall quality level experienced by all the UEs of all slices.

The virtual network assigns the resources on a scheduling-frame basis. It defines with $\phi_{u_s}^t$, ϕ_s^t and Φ^t the quality level in a generic scheduling frame t . Accordingly, we can also define the time-average quality level values as follows:

$$E[\phi_{u_s}] = \frac{1}{T} \sum_{t=1}^T \phi_{u_s}^t \quad (5.4)$$

$$E[\phi_s] = \frac{1}{T} \sum_{t=1}^T \phi_s^t \quad (5.5)$$

$$E[\Phi] = \frac{1}{T} \sum_{t=1}^T \Phi^t \quad (5.6)$$

where T is the overall number of considered scheduling frames.

5.3.3 Physical resources

The physical resources refer to the radio resources available in the virtual network. For the sake of simplicity, it refers to the downlink channel of one macro-cell. The total available bandwidth is denoted by B MHz. The set $\mathbb{M} = \{1, 2, \dots, M\}$ represents the available sub-channels, where the bandwidth of the generic sub-channel m is $b_m = \frac{B}{M}$. The total transmit power P^{TOT} is uniformly allocated to each sub-channel, i.e., $p_m = \frac{P}{M}$.

When assigning the physical resources, the channel conditions of the UEs have been considered. It assumes that channel condition is determined by transmission path loss and shadowing components [2]. The path loss is

5.3. System Model

defined in Table 5.2 and the shadowing fading path loss is assumed to be a Gaussian random variable with zero mean and σ standard deviation equal to $8dB$ [2]. Therefore, the path loss is based on the distance value d_{u_s} between a generic UE and the macro-cell, which is given by Equation (5.7).

$$PL(d_{u_s}) = 128.1 + 37.6 \log_{10}(d_{u_s}) + \log_{10}(X_{u_s}) \quad (5.7)$$

where X_{u_s} is the log-normal shadow fading path loss of the UE [2].

It has also been assumed that the macro-cell receives perfect channel gain information from all UEs belong to different service slices, where h_{m,u_s} is the sub-channel gain for the UE u within slice s and can be defined as $h_{m,u_s} = 10^{-PL(d_{u_s})/10}$ [2]. The data rate of the UE with slice s , denoted with r_{u_s} can be described in Equation (5.8) [133].

$$r_{u_s} = \sum_{m \in \mathbb{M}} \varphi_{m,u_s} b_m \left(1 + \frac{p_m |h_{m,u_s}|^2}{N_0 b_m} \right) \quad (5.8)$$

where N_0 is the noise spectral density and φ_{m,u_s} is the situation of the UE u_s which has been defined as Equation (5.9).

$$\varphi_{m,u_s} = \begin{cases} 1 & \text{if sub-channel } m \text{ is assigned to } u_s \\ 0 & \text{otherwise} \end{cases} \quad (5.9)$$

$\varphi_{m,U_s} = 1$ means that the UE associates with the slice s over the sub-channel m ; otherwise $\varphi_{m,U_s} = 0$.

The flow of admission control mechanism has been described in Fig 5.3.

5.4. Two-tier Admission Control and Resource Allocation

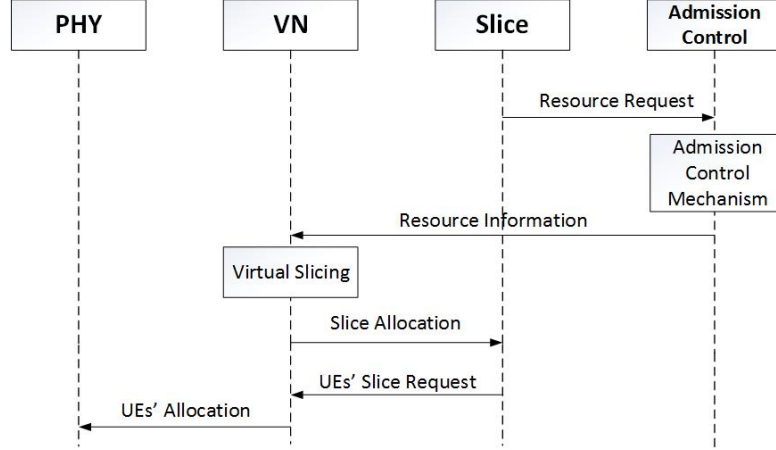


Figure 5.3: Flow of Admission Control

5.4 Two-tier Admission Control and Resource Allocation

In this section, the proposed approach for two-tier admission control and resource allocation will be described.

5.4.1 Admission Control Strategy

The admission control mechanisms aim to allocate service slices to different UEs for the purpose of maximising their quality level while improving the network resource utilisation. It also considers three different kinds of service slices which are TV streaming, car management and Web browsing. TV streaming is a type of service slice with intensive throughput consideration. Moreover, in the 5G mobile networks, TV streaming may have different slice priority values according to their resource requirements. Car management is a type of service slice with lower throughput consideration, while web browsing is a type of service slice with intensive time and lowest throughput consideration. Each kind of service slices contains several UEs sharing their resources simultaneously.

5.4. Two-tier Admission Control and Resource Allocation

First, the admission control mechanism without considering heuristic has been designed in Algorithm 4. In order to improve the network performance compared with the existing 4G networks, a QCI priority based admission control mechanism for service slices is proposed. We consider inter priority order of different slices and intra priority order of different UEs, both of them can help the admission control and slice allocation part to improve the efficiency of network resource utilisation while guaranteeing UEs' QoS requirements in maximisation. It only considers received data rate as its performance parameter which can be extended in the future work.

Then, the heuristic-based prioritised admission control mechanism has been designed in Algorithm 5. This mechanism can be used to deal with the arrivals of new slices or users and provides a global optimisation of the resources allocated to service slices. At each step of the heuristic algorithm, the current solution will be replaced by a new solution with a certain probability. The probability depends on differences between the current solution and randomly generated neighbour solution [107]. For the sake of simplicity, Algorithm 5 refers to the admission control of UEs belonging to the same slice. The steps of the proposed admission control mechanism can be used for admission control of new slices, by easily adapting the parameters under consideration. When the UE enters the network, by considering the quality level of the users in the same slice, it can derive an acceptance probability of the just entered user in the virtual network by considering the constraints in terms of intra-slice priority as well as the quality level of served UEs. In the proposed admission control mechanism, a UE is accepted if the available resources are sufficient to guarantee to satisfy at least the requirement on the minimum data rate. The set of accepted users is thus provided as input to the resource allocation procedure.

5.4. Two-tier Admission Control and Resource Allocation

Algorithm 4: VIRTUALIZED SERVICE RESOURCE SLICING AND AD-
MISSION CONTROL ALGORITHM

Data:

\mathbb{S}_s : slice s ;

\mathbb{M}_m : sub-channel m ;

\mathbb{U}_s : UE associates with slice s ;

Φ_{U_s} : Quality value of U_s ;

k' : the new entering UE $k' \in \mathbb{U}_s$;

$\tilde{\Phi}_{u'_s}$: Quality value of the new entering UE k' ;

```
for  $t := 1$  to  $T$  do
  for  $s := 1$  to  $S$  do
    for  $u := 1$  to  $U$  do
      Calculate  $\Phi_{U_s}$ ;
      find  $X_i = \max\{\Phi_{U_s}\}$ ;
      find  $X_j = \min\{\Phi_{U_s}\}$ ;
      Calculate new quality value of  $k' \in \mathbb{U}_s$ :  $\Phi_{u'_s}$ ;
      if  $E[\Phi_{u'_s}] < \Phi_{u'_s}$  then
        Inject UE  $k'$ ;
        Check inter priority order;
        if the priority order are the same then
           $X_i$  will be replaced by the new UE;
        else
           $X_j$  will be replaced by the new UE;
        end
      end
    end
  end
end
```

5.4. Two-tier Admission Control and Resource Allocation

Algorithm 5: HEURISTIC BASED ADMISSION CONTROL ALGORITHM OF NEW USERS

```

for  $t := 1$  to  $T$  do
  for  $s := 1$  to  $S$  do
    for  $u := 1$  to  $U$  do
      for  $m := 1$  to  $M$  do
        Calculate  $\phi_{u_s} \forall u_s \in \mathbb{U}_s$ ;
        find UE  $x_s$  with the max quality level;
        find UE  $j_s$  with the max quality level;
        while an UE  $u'_s \in \mathbb{U}_s$  enters the network do
          Calculate the new quality value of  $u'_s$ :  $\phi_{u'_s}$ ;
          Then, find the neighbor quality level value of  $u'_s$ :
             $\hat{\phi}_{u'_s}$ ;
          if  $\hat{\phi}_{u'_s} - \phi_{u'_s} > 0$  then
            if  $E[\phi_{u'_s}] < \phi_{u'_s}$  then
              Inject UE  $u'_s$ ;
              check priority order;
              if the priority order are the same then
                 $x_s$  will be replaced by the new UE; else
                   $j_s$  will be replaced by the new UE;
                end
              end
            else
              Do not admit UE  $u'_s$ ;
            end
          end
        else
          generate accept probability  $p = \frac{-\Delta quality}{F}$ ;
          then, the new UE will be rejected based on
            the probability  $p$ ;
          end
        end
      end
    end
  end
end

```

5.4.2 Resource Allocation

The overall problem under consideration during the resource allocation step is the maximisation of the quality level of UEs, by simultaneously considering the inter- and intra-slice priority. This problem can be formulated as in Equation (5.10).

P1:

$$\text{maximize } \sum_{s \in \mathbb{S}} \left[\sum_{u_s \in \mathbb{U}_s} \left(\frac{r_{u_s}}{R_s^{\max}} \right)^{\mu_{u_s}} \right]^{\rho_s} \quad (5.10)$$

$$\text{s.t. } \sum_{m \in \mathbb{M}} \sum_{s \in \mathbb{S}} \sum_{u_s \in \mathbb{U}_s} \varphi_{m,u_s} b_m \leq B, \quad (5.10a)$$

$$R_s^{\min} \leq r_{u_s} \leq R_s^{\max}, \quad (5.10b)$$

$$\sum_{m \in \mathbb{M}} \sum_{U_s \in \mathbb{U}_s} \alpha_{m,U_s} = 1, \varphi_{m,U_s} \in \{0, 1\}, \quad (5.10c)$$

where, constraint (5.10a) indicates that the amount of allocated sub-channels cannot overcome the maximum available bandwidth; this constraint implicitly refers to the orthogonality of assigned resources, too. Constraint (5.10b) indicates that the received data rate by UE u_s is restricted by the requirements of the associated slice s . It is worth noting that, in Equation (5.10), the quality level is a number lower or equal than 1; as a consequence, the higher the priority of a slice, the lower the value of ρ_s . This happens similarly for the users, i.e, the higher the priority of a user, the lower is the value μ_{u_s} . Constraint (5.10c) is orthogonality constraint.

The resource allocation procedure is performed by considering the physical resources available in the network as well as the channel conditions of the UEs.

Each symbol of parameters has been described in Table 5.1.

5.4. Two-tier Admission Control and Resource Allocation

Table 5.1: Notations in Chapter 5

Symbol	Define
b_m	bandwidth of sub-channel
p_m	transmit power of sub-channel
$PL(d)$	path loss
X_{U_s}	log-normal shadow fading
h_{m,U_s}	sub-channel gain of the UE with slice s
r_{U_s}	data rate of UE with slice s
ϕ_{m,U_s}	situation of UE with slice s over the sub-channel
N_0	noise spectral density
$R_{U_s}^{min}$	minimum required data rate by UE with slice s
$R_{U_s}^{max}$	maximum required data rate by UE with slice s
$R_{U_s}^{max}$	max requirments of UE with slice s
μ_{U_s}	priority order over UE in slice s $\sum_{U_s \in \mathbb{U}_s} \mu_{U_s} = 1$
ρ_s	priority order between different slice s $\sum_{s \in \mathbb{S}} \rho_s = 1$

5.5. Performance Investigation

Table 5.2: Main Network Parameters

Parameter	Value
Number of cells	one macrocell
LTE bandwidth	5 MHz
UEs distribution	uniform
Overall number of UEs	100
Overall interval	10s
Noise Spectral Density	-174dBm/Hz

5.5 Performance Investigation

This section provides a performance comparison of the proposal with a legacy 4G resource allocation algorithm. With this aim, a resource allocation algorithm has been implemented where network resources are allocated in order to maximise the overall quality level of users by taking into account their QoS requirements (minimum and maximum data rate) as well as the priority of each user.

In simulations, the arrival rates of UEs have been considered as uniformly distributed within the whole simulation period. The overall number of UEs is uniformly distributed among the considered slices. The priority of UEs within the same slice is randomly generated with the constraint of having a sum equal to 1. In the case of 4G-SA, the priority values of UEs are the same of those considered for 5G-SA and 5G-AC-SA, with the difference that the constraint of having a sum of priorities equal to 1 is extended to all users in the system. The parameters of our simulation model have been listed in Table 5.2 and Table 5.3 [2].

5.5. Performance Investigation

Table 5.3: Application Slice Parameters [2]

	Thr (kbps)	slice priority	4G priority
TV Streaming A	1000-1500	0.1	0.15
TV Streaming B	1000-1500	0.2	0.15
Car Management	400-700	0.2	0.3
Web Browsing	100-300	0.5	0.4

5.5.1 Simulation Cases

- *Case One: 4G slice allocation (4G-SA)* In this case, the benchmark case has been set up, hereinafter named 4G service allocation (4G-SA) which is a single-tier priority algorithm and does not take into consideration the possibility that UEs belong to different slices. The differentiation among the UEs refers only to a different requested service. Only admission control algorithm with received throughput as its performance consideration has been used.
- *Case Two: Slice Allocation with Prioritised Admission Control Mechanism in 5G (5G-SA)* In this case, the slice allocation with prioritised admission control as described in Sec. 5.4.2 has been examined. It considered QCI priority order (inter priority) of different slices and emergency priority order (intra priority) of different UEs. Then, it uses algorithm as explained in Section 4 with received throughput as its consideration as well.
- *Case Three: Heuristic-based Slice Allocation with admission control in 5G (5G-AC-SA)* In this case, prioritised admission control mechanism has been used to combine with the heuristic technology at the same

5.5. Performance Investigation

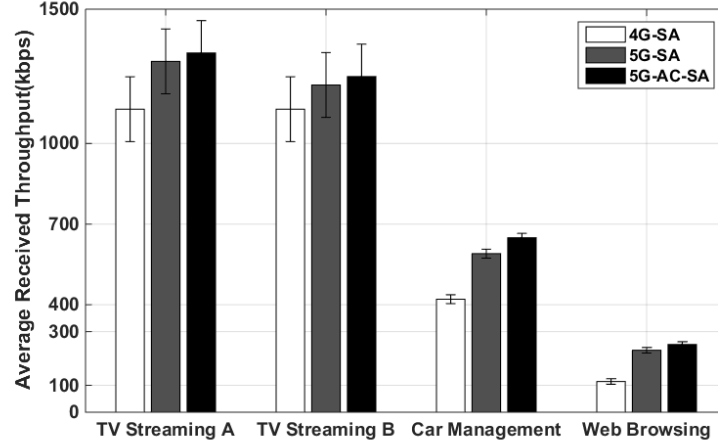


Figure 5.4: Average Cumulative Received Throughput

time. It takes into consideration the admission control procedure in Sec. 5.4.1, which is performed as a first step before the resource allocation. The reason behind this choice is to highlight the impact of the admission control in the management of network resources. Heuristic based admission control mechanism can be used improve network performance from the point the whole system. The quality values of neighbour UEs will not be affected when the current UE has been allocated a higher quality value. Therefore, the network performance can be improved efficiently.

5.5.2 Result Analysis

Fig. 5.4 indicates the average throughput for UEs from different slices. It can be noticed that 4G-SA allocates the minimum data rate to services with low requested data rate (i.e., web browsing and car management) in order to provide higher data rate to the TV streaming services. The reason behind this behaviour is that, to maximise the quality values, 4G systems prefers to boost the performance of users with higher QoS requirements. it is worth noticing that, with 4G-SA, users belonging to TV streaming A and B

5.5. Performance Investigation

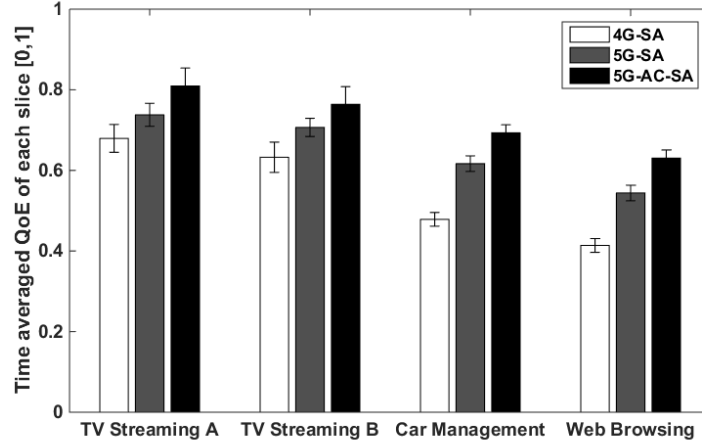


Figure 5.5: Average Quality Levels of Different Slices

experience the same throughput, although these such services belong to two different slices with different priorities. This strongly underlines that 4G-SA is not able to guarantee prioritisation on a slice-basis, but only on a service-basis. The proposed approaches, meaningfully increase the data rate for all the service slices compared to 4G-SA. It can be observed that 5G-SA-AC guarantees higher throughput compared to 5G-SA, by thus highlighting the importance of the proposed admission control in achieving better utilisation of spectrum resources. Finally, it can be noted that both 5G-SA and 5G-SA-AC introduces differentiation in the throughput of slices TV streaming A and B, as they are designed to take into consideration the inter-slice priority.

Fig. 5.5 indicates the average quality level of different slices. It can be noticed that 4G-SA provides the lowest quality level of different slices. In addition, it can be noted that although 4G-SA guarantees the same throughput to TV streaming A and TV streaming B, the quality level of TV streaming B is lower than that of TV streaming A. Because TV streaming B has a lower priority level compared to TV streaming A. When focusing on 5G-SA and 5G-AC-SA, it can be observed that they substantial increase the quality level compared to 4G-SA and they offer a better fairness in the quality level

5.5. Performance Investigation

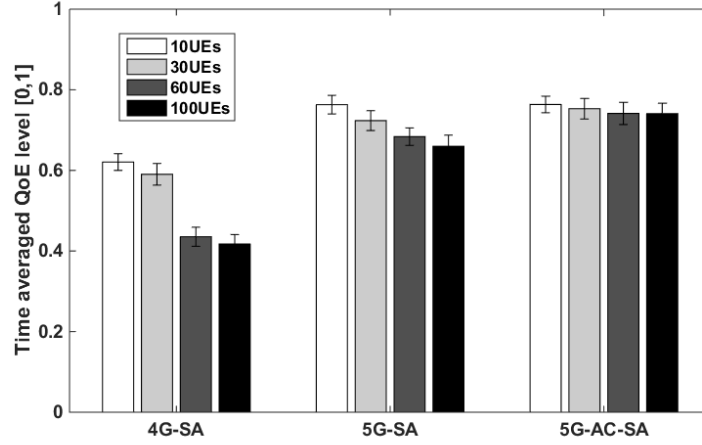


Figure 5.6: Time Averaged total Quality Level with Different Number of UEs

experienced by the users of different slices. Indeed, when focusing on the quality values of TV streaming A and web browsing, it can be observed that 5G-SA and 5G-AC-SA guarantee a lower difference in the quality level of these slices compared to 4G-SA. It means that the proposed approaches are able to guarantee a better fairness compared to 4G-SA as well.

Fig. 5.6 indicates the time averaged total quality level by considering a varying number of UEs per slice. From this figure, it can be observed that the overall quality level for 4G-SA decreases as the number of UEs increases. This is due to the fact that when the number of UEs increases, this algorithm tries to increase the overall data rate of the system by allocating resources to the services with higher data rates (as shown in Fig. 5.6). From a global point of view, this involves a quality level reduction as 4G-SA does not consider inter-service priority. The proposed approaches, on the contrary, are based on the idea of exploiting the inter-slice priority for slice allocation. As a consequence, our approaches do not show a meaningful degradation of the overall quality level when increasing the number of UEs. It is worth noticing the benefits introduced by 5G-AC-SA, that is able to guarantee a lower quality level decrease compared to 5G-SA.

5.5. Performance Investigation

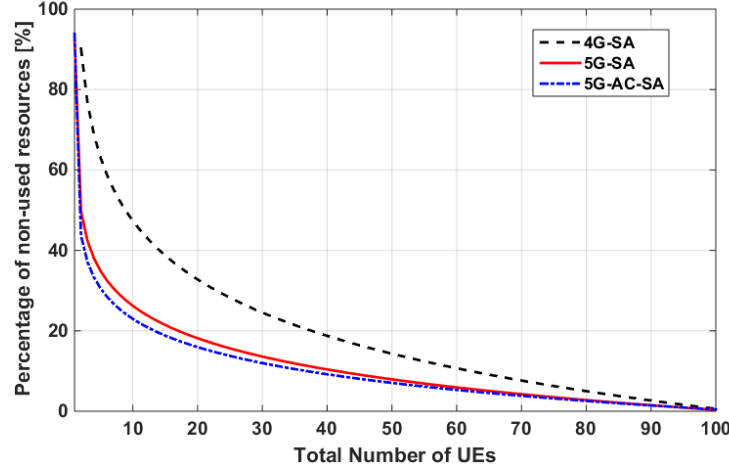


Figure 5.7: Percentage of free resources after the resource allocation step.

Fig. 6.5 indicates the amount of free network resources after the resource allocation. The 4G-SA algorithm maximises the quality level without considering the inter-service priority, i.e., it only considers the priority of users. This means that, to maximise the quality level, the only parameter that meaningfully influences the resource allocation procedure is the QoS requirement. As a consequence, 4G-SA algorithm manages the network resources with the aim to maximise the quality level (as shown in the previous analyses) of users belonging to TV streaming services, as they have higher QoS requirements. Once such UEs are scheduled, from a 4G-SA point of view, the allocation of additional resources to other services does not introduce any meaningful quality level increase, and thus the algorithm stops with the side effect of not assigning a portion of network resources. On the contrary, the proposed approaches consider the inter-slice priority. This means that all UEs in the same slice are grouped together in order to consider the whole priority of the slice (please refer to equation (5.3)). As a consequence, to increase the overall quality level, 5G-Sa and 5G-AC-SA allocate resources also to slices with low priority and this means a better exploitation of the available network resources

5.6 Conclusions

In this chapter, a novel approach for resource allocation in the 5G networks with network slicing has been presented. The proposed approach is a heuristic based prioritised admission control mechanism that takes into consideration both the inter- and the intra-slice priority and performs the resource allocation. Accordingly, in order to meet the QoS requirements dictated by the service slice. The proposed approach increases the quality level experienced by mobile UEs as well as allows a better management of network resources.

Chapter 6

An Auction based Network Slicing in 5G Mobile Networks

6.1 Introduction

The next-to-come fifth generation (5G) mobile network is expected to open unprecedented business opportunities to telco operators by increasing their market to the business owners and providing not only business-to-customer service (B2C) services but also business-to-business (B2B) and business to business to consumer (B2B2C) services. A large number of vertical industries are foreseen as natural users of 5G beyond the mobile broadband services, including healthcare automotive, smart cities, and industry automation [136]. In order to deliver services to such wide range of industries with diverse requirements, *network slicing* has been introduced in 5G networks, which can be then foreseen as a composition of multiple slices, each one designed with a set of functionalities tailored to serve a specific business [137].

From the network performance point of view, slicing implies that each 5G slice needs to have its own set of allocated resources and this aspect introduces a novelty in the management of network resources in mobile systems.

6.1. Introduction

Indeed, in the previous generations of mobile networks, the resources to be assigned to each application were mainly radio resources [138], while in the 5G network, resources represent both radio and core network [139], by means of computational and storage capabilities in addition to the over-the-air data rate. From the business point of view, however, the issue of pricing a given slice is similar to pricing given spectrum, i.e., there is no clear way to value a given created slice by the operator. Therefore, there is a need for a novel business model by the operators.

In order to consider the fact that slice allocation in 5G means allocating resources throughout the network, 5G network resources as multiple *chunks* has been modelled in this chapter, each one with a different capacity, spread across the whole physical network. This allows to take into consideration the management of resources in the core network in addition to the resources in the radio access. In order to optimise the network revenue by considering network utilisation aspects, this chapter focuses on a resource management strategy based on a novel competitive auction mechanism combined with an optimisation algorithm for network resources allocation. Therefore, objectives are increasing network utilisation, enhancing the satisfaction of requirements of network slices and increasing the incentive for operators by maximising their revenue.

The pricing mechanism considered in this chapter is built based on the assumption that sellers (i.e., network chunks) are assumed to have the knowledge of products (i.e., network resources) for potential buyers (i.e., network slices). Using the Bertrand model for pricing [140, 141], network chunks receive the information from network slices through the auction mechanism while network slices receive their optimised price values. To formulate the problem of price decision, a game theory based auction mechanism has been proposed as an effective method for analysing interactive decision making

6.1. Introduction

[142] in order to consider the interaction between network chunks and network slices. For maximising the network revenue, an auction mechanism which is derived from the Vickrey-Clarke-Groves (VCG) [4, 143] approach has been proposed in this chapter. Therefore, compared with previous researches which mainly focused on price definition, a novel auction mechanism has been designed by considering the amount of network resources the network slices are requesting to the network. Concerning the resource allocation, unlike game theory based resource management mechanisms in the more traditional way, the efficiency of resource utilisation has been taken into account with the constraints of maximising the network revenue. Therefore, the Lagrangian based optimisation mechanism has been designed for the purpose of optimising the allocated network resources to network slices without changing their cost values.

The key contributions of this chapter can be summarized as follows:

- Enabling new business model for offering 5G network slices to a wide range of vertical industries, through novel pricing model and auctioning mechanism.
- Presenting a novel augmentation of end-to-end network slices based on both radio and core network resources.
- Formulating a novel resource management model considering points 1 and 2.
- Investigating performance of the discussed models thoroughly, and through simulations.

The remainder of this chapter is organized as follows. Section 6.2 focuses on the business models for 5G network slicing and also review the state of art in resource allocation strategies for network slicing and auction mechanisms.

6.2. Related Works

After elaborating the system model and concept definitions in Section 6.3, the problem formulations, the proposed resource sharing mechanism and pricing mechanism will be described in Section 6.5. Section 6.6 provides numerical results to verify our design objectives and analyses performance observations. Finally, summary of our work is detailed in Section 6.7.

6.2 Related Works

Network slicing drives the business models behind 5G ecosystem by providing an effective way to delivery heterogeneous services of interest for different verticals [144]. As previously discussed, there are three different of business models for network slicing: B2B, B2C, and B2B2C [145]. In B2B model, operators sell customised network resources to enterprises and release full control of end consumers to enterprises. In B2C model, customers purchases customised network resource based on their requirements without considering which operator provides the requested resources. In B2B2C model, operators just provide customised network resources to a broker, and the broker gets more control of the network and engages with end consumer directly. It can be thus noticed that above mentioned business models deal with the allocation of network resources; consequently, the effective revenue as well as the network performance directly depend on the way the operator manages such resources. Nevertheless, resource allocation for network slicing and auction mechanisms to improve network utilisation or revenue have been addressed in literature as two separated contexts. In the remainder of this Section, an overview of the related work from both resource allocation and auction points of view have been proposed and then the novelties of this chapter has been described with respect to the existing works in literature.

From a resource allocation point of view, network slicing is strictly re-

6.2. Related Works

lated to virtualization [146], enabling the management (deployment, placement, moving, etc.) of network functionalities across the network. Several solutions for efficiently supporting network resource virtualization [129] and resource allocation by using auction approaches [4] have already been proposed. They have been designed to improve the quality level of mobile users (thus focusing on an end-user point of view and not from a slice point of view) and network utilisation. In this section, a briefly overview of the most relevant studies will be described. Focusing from a network resource slicing point of view, a resource allocation strategy of virtualized resources for Long Term Evolution (LTE) networks has been proposed in [133]. This work proposed a slicing scheme to allocate virtualized radio resources, i.e., resource blocks, to different service providers (SPs) in order to maximise the radio utilisation. The proposed scheme was dynamic and flexible for addressing arbitrary fairness requirements of different SPs. Similarly, [134] proposed a framework of wireless resource virtualization in LTE networks to allow radio network resources to be shared among mobile network operators. An iterative algorithm has been proposed to solve the Binary Integer Programming (BIP) problem with less computational overhead. Finally, in the previous work [147], the topic of slice association and resource allocation for mobile users have been mainly considered with the aim of increasing the quality level the users.

From an auction point of view, game theory based resource auction mechanism have been widely investigated in existing works [148]. Nash equilibrium was considered as the solution for solving the problem of spectrum sharing in cognitive radio networks in [149]. Stackelberg game mechanism has been formulated for power control in wireless networks to maximise the capacity [150] [151], while a non-cooperative game based power control algorithm has been proposed in [152] together with a base station association

6.2. Related Works

scheme for heterogeneous networks. In [153], an auction mechanism has been proposed to maximise the expected revenue of sellers. An auction-based scheme which can be used to develop a synchronous algorithm for solving the optimisation problem of resource allocation has been proposed in [154]. The game theory based network virtualization framework has been described in [155] and an auction mechanism has been used for pricing the instantaneous rate consumption. Compared with the other game theory mechanisms, the auction mechanism was widely applied [154] [156] in the situation of competitive resource allocation. Meanwhile, effective allocation of network resources can improve the revenue of both users and networks. In [140], a competitive pricing model has been formulated in a dynamic spectrum access where a few primary services offer spectrum access opportunities to a secondary service. In [157], a double-auction-based algorithm for inter-cloud virtual machine trading has been designed. The proposed algorithm was strategy-proof and used for the sell and purchase of available virtual machines across cloud boundaries over time. The proposed algorithm was proven to maximize individual profit for each cloud over the long run of the system. In [158], an auction based online mechanism for virtual machine provisioning, allocation, and pricing in clouds that considers several types of resources has been proposed. The proposed mechanism calculated the allocation and payments as users arrive at the system with their requirements, and aimed to maximize users' utilities.

The above considered resource allocation mechanisms were mainly considering the radio segment (e.g., power allocation, radio resources) without taking into account the other resources in the mobile core networks such as storage and computational resources. In addition, the competitive price mechanism in the existing research did not consider the global resource utilisation as well as priority values (i.e., network resources spread across the edge

6.3. System Model

of the network are scarcer than the ones in a central cloud and are more subject to overload) of different network resources. From the discussion above, the relationship between resource allocation and operator revenue becomes one of the key aspects when designing a business-driven network slice allocation. With this aim, this chapter extends the approaches in literature by proposing an auction business model for the operators which takes into account the availability of different types of network resources across the network (i.e., both radio and core network segments). The proposed auction mechanism is modelled by considering the total revenue for the operator and the total amount of assigned resources.

6.3 System Model

In this section, system model will be described in detail as in Fig 6.1 and further elaboration are different part of the model are given in the following subsections. The network architecture is divided into two parts, i.e., the radio access layer (handling over-the-air resources) and the mobile core layer (handling computational and storage resources and consisting of two parts as central cloud and an operator cloud located at the edge).

6.3.1 Network Slices

The set of network slices requesting is denoted by $\mathbb{K} = \{1, 2, \dots, K\}$. The concept of slice priority has been introduced in order to consider the fact that some slices could have different importance from an operator point of view (e.g., a slice related to emergency services or a slice of a vertical which is a premium business partner of the operator). For each slice k , the priority has been defined in the range of $\rho_k = \{1, 2, \dots, A\}$, where A indicates the maximum priority level in the system model. The higher ρ_k , the higher the

6.3. System Model

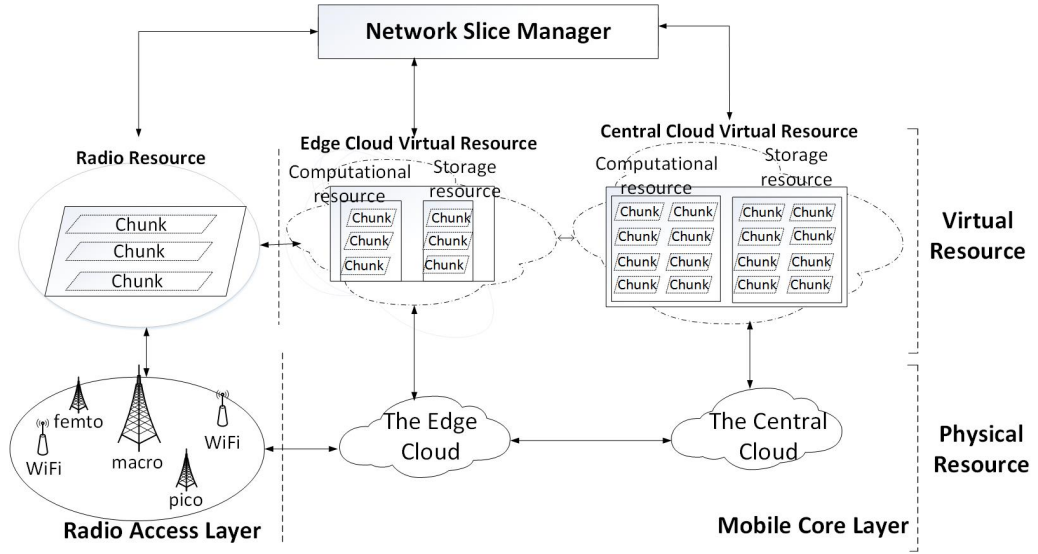


Figure 6.1: System model

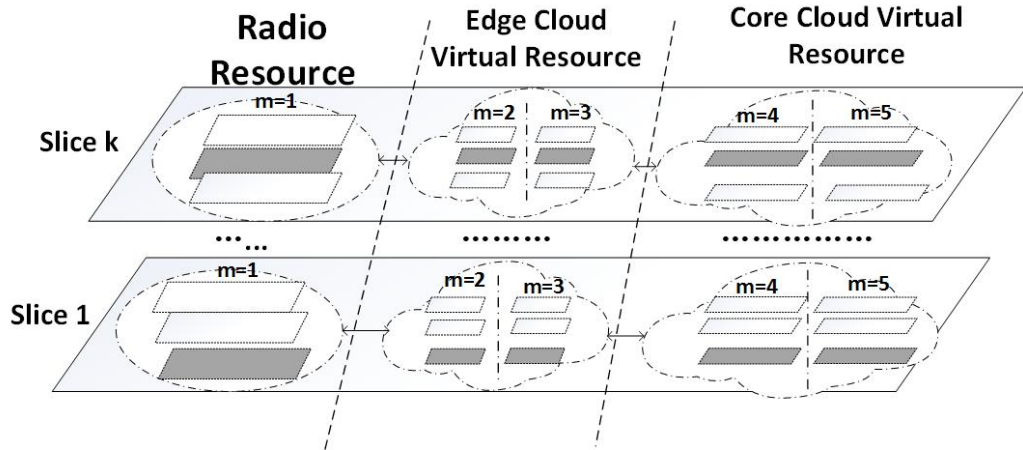


Figure 6.2: Slice Model

priority for the given slice, k .

6.3.2 Network Resources

In the system model, as shown in Fig 6.1, the available network resources are modelled as diverse types of network chunks, and the total number of available network chunks is denoted by $\mathbb{M} = \{1, 2, \dots, M\}$. From a radio access point of view, different chunks can be seen as different radio access technologies (RATs) providing wireless resource to the network slices. From a core network point of view, the chunks offer resources in terms of compu-

6.3. System Model

tational capability (i.e., CPU) and resource storage capacity (i.e., RAM) in two different locations, i.e., a central cloud and an operator cloud close to the edge.

The resources of each network chunk m are allocated to different slices, as in Fig 6.2. The capacity of each network chunk m is denoted by η_m , where $m \in \mathbb{M}$. The amount of resources assigned to slice k from resource chunk m is denoted by $\sigma_{k,m}$, where $m \in \mathbb{M}$ and $k \in \mathbb{K}$, respectively. Therefore, the ratio of network resource allocation that network slice k receives from the network chunk m can be explained as $\frac{\sigma_{k,m}}{\eta_m}$. Moreover, in the real network environment, different network chunks will have different amount of network resources which have been described as “weight values” denoted by β_m . This choice allows us to abstract the effective capacity of each network chunk and to focus only on the portion of resources each slice requires from each chunk.

The network slice manager, in charge of running the auction process, price definition and allocating network resources to slices, is assumed to be an impartial entity (i.e. third party). In this case, cost values of different network chunks’ per unit have been described as $\Omega_m = \{\omega_m | m \in \mathbb{M}\}$, and $C_m = \{c_m | m \in \mathbb{M}\}$ indicates the selling price per unit of the network chunk m . The network slice manager will make auction-based resource allocation decisions for network slicing. It is assumed that each network slice is able to provide the slice manager with a request in terms of needed resources. Therefore, network chunks with the best pay-off values generated based on Equation (6.9), will be allocated to the slices (details can be seen in Fig 6.3).

6.3.3 Slice Requirements

As shown in Fig 6.2, each network slice in the architecture requires resources in different types of network chunks, which are five in our model: over-the-air resources, the computational and the storage resources from the operator

6.4. Operators Revenue and Auction Model

cloud at the edge, and the computational and the storage resources from the central cloud. The radio resource required by the network slice k from the network chunk m is denoted by $\mathbb{R}_{k,m}$. The value of $r_{k,m}$ represents the minimum resource requirements by network slice k from the network chunk m . The value of $R_{k,m}$ represents the maximum resource requirements by network slice k from the network chunk m ; such values are assumed to be an indication of the needs of the slices.

6.4 Operators Revenue and Auction Model

In the above explained system model, competitive resource allocation mechanism involves three aspects: *a)* the price competition model, where different operators compete for achieving selling prices to maximize their revenue; *b)* an auction based resource allocation mechanism for network slicing, where different slices are competing in order to maximize the network revenue; *c)* an optimised resource allocation for all required slices.

Before determining resource allocation for maximising the network revenue and optimising the amount of resources allocated to network slices, the network slice manager should determine the selling price c_m of the network chunk m by using auction mechanism. Then, competitive auction mechanism of network slicing and optimisation algorithm for allocating resource to all slices are performed.

6.4.1 Price Competition: Bertrand model

As mentioned earlier, the price competition among different network chunks is defined based on Bertrand model, where limited network chunks can compete with each other in order to achieve the optimal profit by controlling the price of each network chunk. The assumption is that both competing parties

6.4. Operators Revenue and Auction Model

have the same constant unit cost of production (i.e. unit cost of network chunks), so that marginal and average costs are the same and equal to the competitive price.

Such price competition can be applied to analyse and obtain equilibrium price strategies over the 5G mobile networks. The system model contains several network chunks, and these chunks can be sold to network slices. The price strategies are determined by the network slice manager based on utility value and capacity of network chunks as well as the requested resource by network slices.

The profit maximisation problem of the network chunks while serving all slices can be formulated as follows,

$$\text{P1: maximize } \Theta(c_m, \sigma_{k,m})$$

The revenue function of $\Theta(c_m, \sigma_{k,m})$ is defined in Equation (6.1).

$$\Theta(c_m, \sigma_{k,m}) = \sum_{k=1}^K \sigma_{k,m} c_m - \omega_m \eta_m, \quad (6.1)$$

where ω_m is the cost value per unit of the virtual network chunk m . The value of ω_m has been defined as $\omega_m = \frac{\eta_m}{\sum_{k=1}^K \sigma_{k,m}}$, while the value of $\omega_m \eta_m$ is fixed. The maximum value of Θ is equal to zero.

The revenue value per unit of the network chunk m is given by,

$$\theta_m(c_m, \sigma_{k,m}) = \sum_{k=1}^K \sigma_{k,m} c_m - \omega_m. \quad (6.2)$$

The partial differential of the revenue of virtual network chunk unit is given by,

$$\begin{aligned} \frac{\partial \theta_m(\eta_m)}{\partial c_m} &= \frac{\partial}{\partial c_m} \left(\sum_{k=1}^K \sigma_{k,m} c_m - \omega_m \right) \\ &= \sum_{k=1}^K \sigma_{k,m} + c_m \frac{\partial \sum_{k=1}^K \sigma_{k,m}}{\partial c_m} - \omega_m \frac{\partial \eta_m}{\partial c_m}. \end{aligned} \quad (6.3)$$

6.4. Operators Revenue and Auction Model

In order to maximize total revenue for each network chunk, revenue per unit of network chunk is maximized. Therefore, $\frac{\partial \theta_m(\eta_m)}{\partial c_m} = 0$, hence Equation (6.3) is transformed to Equations (6.4), and (6.5).

$$\sum_{k=1}^K \sigma_{k,m} + c_m \frac{\partial \sum_{k=1}^K \sigma_{k,m}}{\partial c_m} = \omega_m \frac{\partial \eta_m}{\partial c_m}, \quad (6.4)$$

$$\frac{\sum_{k=1}^K \sigma_{k,m}}{\frac{\partial \eta_m}{\partial c_m}} + c_m \frac{\partial \sum_{k=1}^K \sigma_{k,m}}{\partial \eta_m} = \omega_m. \quad (6.5)$$

In Equations (6.4) and (6.5), $\sum_{k=1}^K \sigma_{k,m} = \eta_m$, assuming resource of the network chunk m has been fully allocated to network slices. Therefore,

$$\frac{c_m}{\frac{\partial \eta_m}{\partial c_m} \frac{c_m}{\eta_m}} + c_m = \omega_m \quad (6.6)$$

$$c_m \left(\frac{1}{\frac{\partial \eta_m}{\partial c_m} \frac{c_m}{\eta_m}} + 1 \right) = \omega_m \quad (6.7)$$

Let $\varepsilon = \frac{\partial \eta_m}{\partial c_m} \frac{c_m}{\eta_m}$, where ε is a small positive value. Therefore, the value of c_m based on the Bertrand Nash equilibrium can be given by $c_m = \frac{\varepsilon}{\varepsilon+1} \omega_m$. Based on the result of Bertrand Nash Equilibrium mechanism, we define a proposition which can be described as follows:

Proposition 1: Based on Bertrand model, equilibrium price values of network chunks per unit is given by $c_m = \frac{\varepsilon}{\varepsilon+1} \omega_m$. This is the selling price per unit of the network chunk m that will be charged to the network slice k while no unused resource in the edge cloud is generated.

Proof 1: If $c_m = \frac{\varepsilon}{\varepsilon+1} \omega_m$, the value of Θ is given by Equation (6.8), where ε is a small positive value (explained earlier) and $\sum_{k=1}^K \sigma_{k,m} = \eta_m$, hence Θ equals to 0.

$$\Theta(C_m, \sigma_{k,m}) = \sum_{k=1}^K \sigma_{k,m} \frac{\varepsilon \omega_m(\eta_m)}{\varepsilon + 1} - \omega_m \eta_m \quad (6.8)$$

6.4. Operators Revenue and Auction Model

In Equation (6.8), the total profit value of allocated resources is equal to the profit value of the total network resource capacity. In this case, there is no incentive for competitors to deviate from their optimized cost values. That is because if one of the competitors chooses to improve the selling price, the selling price will be increased as well by allocating its resource. If one of the competitors chooses to decrease the selling price, it will not have enough resource to meet the increased demands. Therefore, the selling price generated from the competitive mechanism is the optimal value for maximising the function Θ .

Given c_m is the price per unit of the network chunk m , we can also define the network slice paid value according to their priority levels as,

$$p_{k,m} = \rho_k c_m \beta_m \quad (6.9)$$

where $p_{k,m}$ is the price paid by network slice k to the network chunk m , and $c_m \beta_m$ indicates the price of network chunk m .

6.4.2 Auction Mechanism for Network Slicing

In this section, properties of the auction model are detailed. The proposed auction mechanism is based on the VCG mechanism, which has been widely used for pricing in order to improve the efficiency of mobile/wireless resource allocation [154, 156]. According to the system model used in this paper, network slices are players and the aim of auction mechanism is to maximise network revenue by making strategic decisions for all network chunks, i.e., by looking at the auction mechanism from an overall point of view instead of focusing on only one network chunk. Furthermore, the auction mechanism satisfies the social efficiency (social efficiency is guaranteed of all participants' requests are optimised over all feasible allocation mechanisms.).

6.4. Operators Revenue and Auction Model

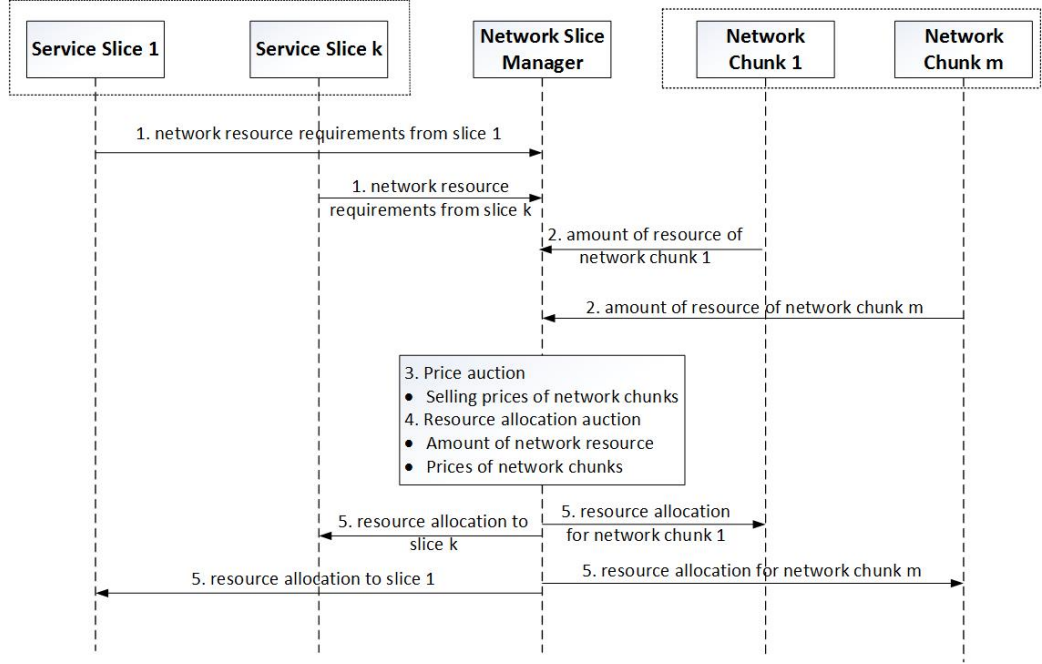


Figure 6.3: Auction Flow Chart. Steps of our proposed auction model can be mainly described as: Network slices send a resource request to the slice manager which runs the auction mechanism considering the revenue and the network utilisation of all the chunks of the network. Afterwards, slices will run auction mechanism in order to access their optimised resources of all network chunks.

The network resource of each chunk allocated to service slices are determined by their own resource requirements and the impact on other network slices. Therefore, the proposed auction based network slicing mechanism can be depicted as follows.

- In the first step, the optimal amount of network chunks for network slice k is chosen based on its resource requirements. The generated network revenue because of network slice k is denoted by $U_k(\sigma_{k,m}, p_{k,m}, \omega_m)$ as in equation (6.10).

$$U_k = \sum_{m=1}^M p_{k,m} \sigma_{k,m} - \sum_{m=1}^M \omega_m \quad (6.10)$$

- In the next step, the network revenue is calculated after removing slice k from the network. It can be described as $U_{\mathbb{K}/\{k\}}(\sigma_{\mathbb{K}/\{k\},m}, p_{\mathbb{K}/\{k\},m}, \omega_m)$.

6.4. Operators Revenue and Auction Model

In addition to allocating network resources to slice k based on its resource requirements, the VCG mechanism force the network slice manager to consider profit values of other $K - 1$ network slices. Given the objective of maximising total network revenue, function of $U_{\mathbb{K}/\{k\}}$ has been defined in equation (6.11).

$$U_{\mathbb{K}/\{k\}} = \sum_{\mathbb{K}/\{k\}} \sum_{m=1}^M p_{\mathbb{K}/\{k\},m} \sigma_{\mathbb{K}/\{k\},m} - \sum_{m=1}^M \omega_m \quad (6.11)$$

- Finally, the total network revenue is maximized by computing $\sum_{k=1}^K (U_k - U_{\mathbb{K}/\{k\}})$.
- The revenue generated by slice k based on auction mechanism Γ , and the total revenue can be expressed as equations (6.12) and (6.13) consecutively.

$$U_k^\Gamma = U_k - (\max U_{\mathbb{K}/\{k\}}) \quad (6.12)$$

$$U^\Gamma = \sum_{k=1}^K U_k^\Gamma \quad (6.13)$$

- Hence, the optimized network resources allocated to slice k based on the auction mechanism can be expressed as equation (6.14).

$$\sigma_{k,m}^* = \arg U_k^\Gamma \quad (6.14)$$

The idea behind the proposal aiming at maximising the revenue coming while keeping into consideration the amount of resources allocated to network slices, can be expressed with the following proposition.

Proposition 2: In order to allow network slice manager to allocate available network resources using competitive auction mechanisms, while

6.5. Problem Formulation

maximising profit of each network slice k , and maximising total network revenue, all slices should provide their minimum and maximum requirements.

Proof: Based on the above described auction mechanism, from the network slice side, the optimal received resources of each network chunk m can be expressed as $\sigma_{k,m}^*$, and profit of the network is explained in equation (6.15).

$$q_k^* = \frac{\sum_{m=1}^M p_{k,m} \sigma_{k,m}}{\sum_{m=1}^M c_m \eta_m} \quad (6.15)$$

In order to maximize U^Γ , the $p_{k,m} \sigma_{k,m}$ should also be maximized; this shows the relationship between the network revenue and the amount of allocated resources, i.e., $\sigma_{k,m}$.

If all network slices provide the network with their requirements in terms of upper and lower amount of expected resources, all profit values q_k^* can be optimised according to our proposed auction model. So doing, a minimum amount of resources assigned to each slice (this means assigning at least the minimum amount of requested resources to each slice) should be guaranteed and the remaining amount of resources should be assigned with the aim to maximise the total network revenue. Otherwise, if network slices do not provide their resource requirements in terms of upper/lower bounds, the resource allocation task is not able to properly assign the available network resources to the slices; as noted in (6.12) and (6.13), this would decrease the total network revenue.

6.5 Problem Formulation

6.5.1 Problem Formulation

In this section, formulations of the problem of interest are detailed. From an economic point of view, the total network revenue is maximised (i.e., the total

6.5. Problem Formulation

revenue generated from the allocation of resources in all network chunks). From the demand point of view, the allocated network resources throughout the network has been taken into account with the aim of keeping the level of network load, low. Hence, the trade-off between the maximisation of operator's revenue and the minimization of the network load measured in terms of allocated resources is expressed in problem P2.

$$\text{P2: } \arg \max_{\sigma_{k,m}^*} \frac{U^\Gamma}{\sum_{k=1}^K [\sum_{m=1}^M \sigma_{k,m}^*]^{\alpha_k}}$$

$$\text{s.t. } \sum_{k=1}^K \sigma_{k,m}^* \leq \eta_m \quad (6.16a)$$

$$r_{k,m} \leq \sigma_{k,m}^* \leq R_{k,m}, \quad \forall k \in K, \quad \forall m \in M \quad (6.16b)$$

In Equation (6.16), U^Γ is the total network revenue gained from network slice k , using our auction mechanism; $\sigma_{k,m}^*$ is the allocated resources to slice k from the network chunk m and α_k is the normalized priority level for slice k . Constraint (6.16a) indicates that the amount of allocated network resources cannot exceed the maximum available resources for each network chunk, while constraint (6.16b) indicates that the received network resources of slice k should satisfy its own resource requirements.

Jain's fairness index has also been used, denoted by ξ , to analyze distribution of network resources among slices. The Jain's fairness index is widely used in literatures to evaluate the level of fairness achieved by resource allocation algorithm [160, 161]. The Jain's fairness index corresponds to all network resources, $T - \xi$, is given in Equation (6.18). The fairness index among slices in accessing resources of each network chunk, $C - \xi$, can be defined as in Equation (6.17).

6.5. Problem Formulation

$$C - \xi_m = \frac{(\sum_{k=1}^K \sigma_{k,m}^*)^2}{K(\sum_{k=1}^K \sigma_{k,m}^*)^2} \quad (6.17)$$

$$T - \xi = \frac{(\sum_{k=1}^K \sum_{m=1}^M \sigma_{k,m}^*)^2}{KM \sum_{k=1}^K \sum_{m=1}^M (\sigma_{k,m}^*)^2} \quad (6.18)$$

6.5.2 Resource Slicing Mechanism

6.5.2.1 Price Auction Mechanism

In this chapter, Bertrand auctioning model has been used for each set of network chunks in order to receive the selling price of network chunks per unit. The **players** in each step are sets of network chunks. The **strategy** of each player is the price per unit of network resources which are non-negative. The **solution** of this mechanism is based on Bertrand auction mechanism.

According to the network resources required by slice k , the network revenue can be expressed as in Equation (6.13). The price auction mechanism is a strategy where network chunks cannot increase their profit by choosing a different action without affecting others [140]. In this case, the optimised result is obtained by defining strategy for all participants from the overall view of the auction model. Selling price for each network chunk is computed according to the of Bertrand auction mechanism in Section 6.4.1.

6.5.2.2 Resource auction Mechanism

The VCG based resource auction mechanism aiming to maximise the network revenue has been described in Algorithm 6. In the proposed mechanism, the benefit of using VCG mechanism is that it can confirm all the participants provide their truthful requirements and receive the optimal resource based on their demands. If the slice does not provide their truthful resource requirements, the allocated resource $\sigma_{k,m}$ will be the second value from the

6.5. Problem Formulation

resource provided list. Therefore, the network revenue and the quality level values of slices will be less than that optimised.

The auction-based resource allocation mechanism aims to maximise the network revenue and has been described in Algorithm 6. In the proposed mechanism, the benefit of using our auction mechanism is that it can guarantee to all the participants their QoS requirements, i.e., the assignment of, at least, the minimum amount of requested resources for each chunk. Moreover, while satisfying allocated network resources $\sigma_{k,m}$ to all participants, network revenue can be improved by our proposed auction mechanism as well.

6.5.2.3 Optimisation Mechanism

Since the optimisation problem P2 needs to be solved only when network slices are allocated, real-time solution is not required. Therefore, we solve P2 using Lagrangian multipliers, as described as in Equation (6.19).

$$\begin{aligned}
L(\sigma, \lambda, \mu, \phi) = & \frac{U^\Gamma}{\sum_{k=1}^K [\sum_{m=1}^M \sigma_{k,m}^*]^{\alpha_k}} \\
& + \sum_{m \in \mathbb{M}} \lambda_m (\eta_m - \sum_{k=1}^K \sigma_{k,m}^*) \\
& + \sum_{k \in \mathbb{K}} \mu_k (\sum_{m \in \mathbb{M}} R_{k,m} - \sum_{m \in \mathbb{M}} \sigma_{k,m}^*) \\
& + \sum_{k \in \mathbb{K}} \phi_k (\sum_{m=1}^M r_{k,m} - \sum_{m=1}^M \sigma_{k,m}^*).
\end{aligned} \tag{6.19}$$

Based on Equation (6.19), one slice can receive the optimized resource value $\sigma_{k,m}^*$, while the following holds:

$$\begin{cases} \partial L_{\sigma_{k,m}}(\sigma_{k,m}^*, \lambda_m, \mu_k, \phi_k) = 0 \\ \partial L_{\lambda_m}(\sigma_{k,m}^*, \lambda_m, \mu_k, \phi_k) = 0 \\ \partial L_{\mu_k}(\sigma_{k,m}^*, \lambda_m, \mu_k, \phi_k) = 0 \\ \partial L_{\phi_k}(\sigma_{k,m}^*, \lambda_m, \mu_k, \phi_k) = 0 \end{cases} \tag{6.20}$$

6.5. Problem Formulation

Algorithm 6: TWO-STEP AUCTION MECHANISM FOR RESOURCE ALLOCATION

k : number of required network slices;

η_m^{rest} : the rest resource of chunk m ;

for $k := 1$ **to** K **do**

if $R_{k,m} < \eta_m^{rest}$ **then**

 Calculate the network revenue because of slice k based on the

 Equation 6.10 where the value of received resource $\sigma_{k,m}$

 equals to the value of their maximized requirements $R_{k,m}$;

 Then, calculate the network revenue value of the rest slices

$K - 1$ according to the Equation (6.11) where the value of

 received resource $\sigma_{K/\{k\},m}$ equals to the value of their

 maximized requirements $R_{K/\{k\},m}$;

end

if $r_{k,m} < \eta_m^{rest} < R_{k,m}$ **then**

 Calculate the network revenue because of slice k based on the

 Equation (6.10) ;

 Then, calculate the network revenue value of the rest slices

$K - 1$ according to the Equation (6.11);

end

 Calculate the revenue value of slice k based on Equation (6.13).

end

6.6. Performance Investigation

The solution is summarised in Algorithm 7, which explains how allocated resources to slices are optimized according to their requests while network revenue is maximised simultaneously.

Algorithm 7: OPTIMIZED THREE-STEP AUCTION MECHANISM FOR RESOURCE ALLOCATION

```
if  $R_{k,m} < \eta_m^{rest}$  then
    Calculate  $\sigma_{k,m}$  based on Equation (6.14) and (6.20) separately,
    which have been denoted by  $\sigma_{k,m}^*$  and  $\sigma_{k,m}$ , respectively;
    if  $\sigma_{k,m}^* \geq \sigma_{k,m}$  then
        Calculate the network revenue value of slice  $k$  and the network
        revenue of the rest of slices  $K - 1$  as Algorithm 6;
    else
        Calculate the network revenue of the slice  $k$  and the
        network revenue of the rest of slice  $K - 1$  as Algorithm 6
        with received value from Equation (6.20).
    end
end
end
Calculate the revenue value of slice  $k$  based on Equation (6.13).
```

6.6 Performance Investigation

This section provides performance investigations of the proposed network slicing. In the simulations, five different types of network resources have been considered which include the radio network chunk, the edge network and the core network chunks. The radio network chunk ($m = 1$) provides over-the-air resources, while the edge and the core network chunks provide storage and computational resources to network slices ($m = 2, 3, 4$, and 5).

It has been assumed that the amount of network resources in each network

6.6. Performance Investigation

chunk is 100% and the resource required by network slices from each network chunk are a portion of that, i.e. for each chunk, the amount of allocated network resources to each slice is $\sum_{k=1}^K \sigma_{k,m}^* \leq 100\%$. It has been assumed that the computational and the storage resources in the core segment of the network are split between the edge and the central cloud such that 30% of these resources are allocated to the edge cloud while the remaining 70% are located at the central cloud of the core network. In other words, the radio chunk ($m = 1$) contains all the over-the-air network resources, computational and storage capabilities for chunks $m = 2$ and $m = 3$ (i.e., the edge cloud) are the 30% of overall network capabilities, and chunks $m = 4$ and $m = 5$ (i.e., the central cloud) contain the 70% of the overall computational and storage capabilities, respectively, of the core network.

6.6.1 Simulation Scenarios

Three different strategies have been considered for network slicing, as described below.

priced-based network slicing (PB-NS) The priced-based network slicing (PB-NS) scenario has been used as a benchmark in this chapter, and it is implemented by considering the minimization of allocated resources across the network.

Two-step auction mechanism for network slicing (TA-NS) The second scenario is based on the proposed two-step auction mechanism for network slicing (i.e. TA-NS), aimed at maximising the network revenue. This is based on solving problem P2, and the solution is detailed in Algorithm 6, in section 6.5.2.2.

Three-step combination auction mechanism for network slicing (TC-NS) The third scenario is based on the proposed three step combination auction mechanism for network slicing (i.e. TC-NS). In this scenario,

6.6. Performance Investigation

it is still focused on problem P2, but also additionally maximising slices satisfaction, as detailed in Algorithm 7.

6.6.2 Results Analysis

Network Slices with the same Requests: In the first step, it has been assumed that all network slices have the same priority value ($\alpha_k = 5$) and request the same amount of resources (equals to $\frac{\frac{R_{k,m}}{\eta_m} - \frac{r_{k,m}}{\eta_m}}{2}$, where $\frac{R_{k,m}}{\eta_m} = 0.05$, $\frac{r_{k,m}}{\eta_m} = 0.01$) from each network chunk.

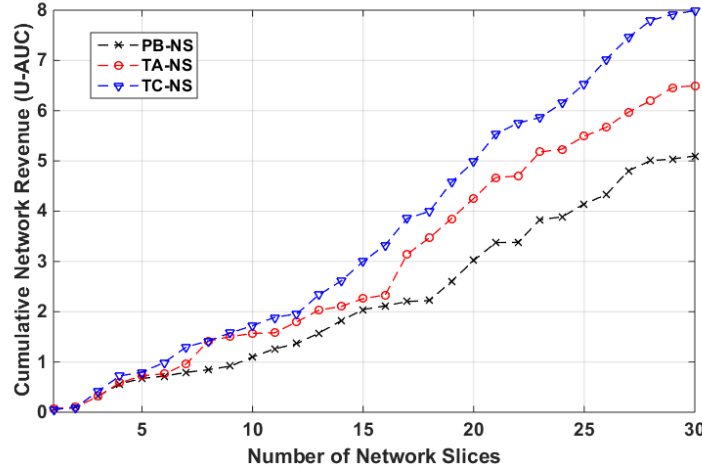
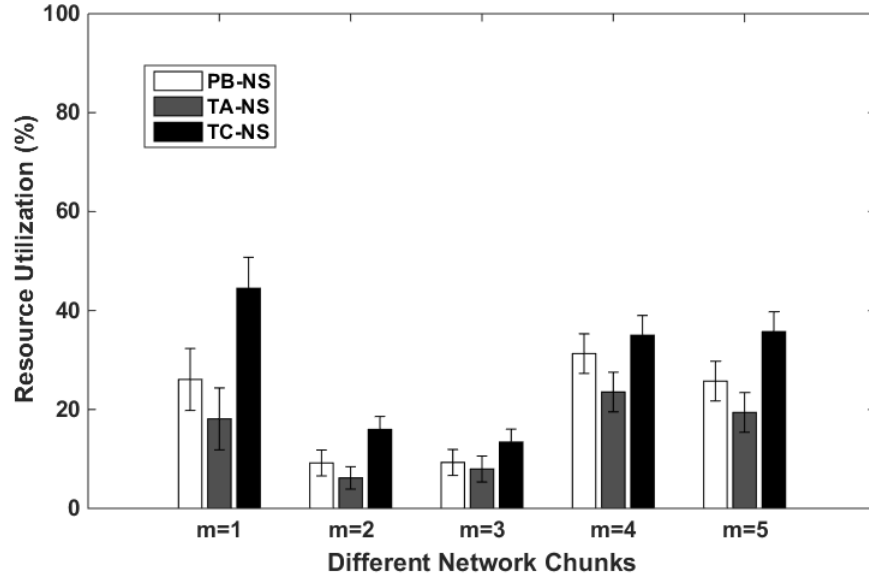


Figure 6.4: Cumulative Network Revenue

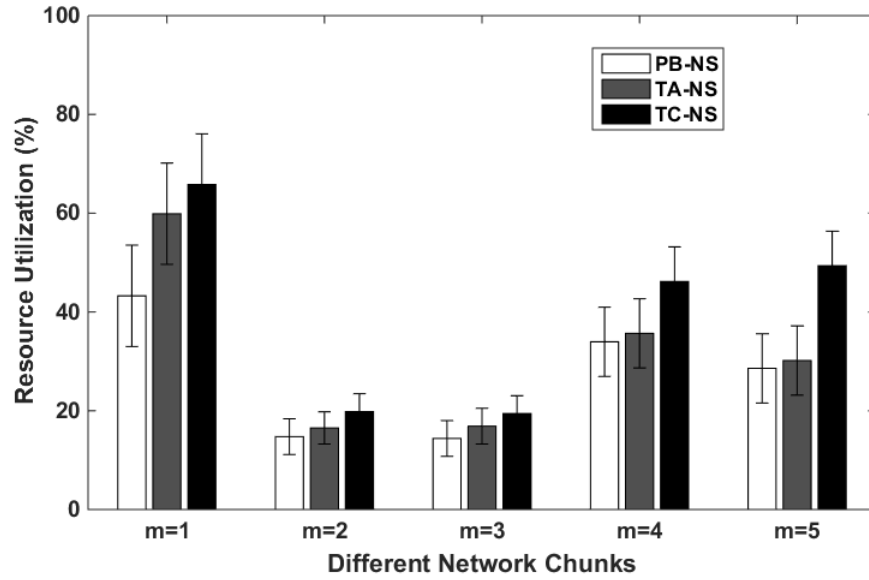
Fig. 6.4 indicates the cumulative network revenue versus a total number of network slices. It can be observed that the highest aggregated network revenue is obtained with the TC-NS strategy. This result is expected based on the definition of the TC-NS scenario.

Fig. 6.5 depicts resource utilisation of different network chunks. It can be seen that the TC-NS scenario allocates the highest amount of network resources to the slices, i.e. has the highest resource utilisation. This is mainly due to the fact that goal of TC-NS is maximising allocated network resources to the slices. In the Focusing on TA-NS, the lowest resource utilisation can be observed in the case with 10 slices, since the aim of TA-NS is mainly

6.6. Performance Investigation



(a) 10 Network slices



(b) 30 Network slices

Figure 6.5: Analysis of Resource Utilisation

6.6. Performance Investigation

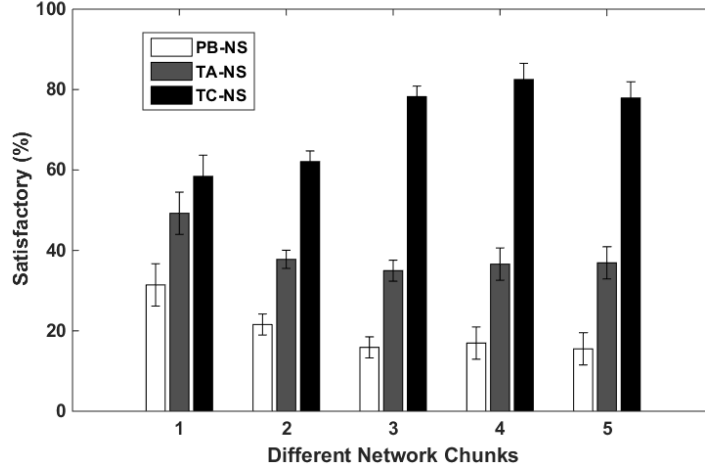


Figure 6.6: Average level of satisfaction from each network chunks

maximising the network revenue. Hence, utilisation in PB-NS is higher than the TA-NS, as shown in Fig. 6.5(a). However, in Fig. 6.5(b), slightly higher utilisation can be observed in TA-NS comparing with PB-NS. This is mainly because TA-NS can support higher number of network slices more efficiently compared with that in PB-NS. In PB-NS, network resources have been allocated only based on price auction mechanism without considering their capacity requirements. Further observation from the two plots of Fig. 6.5 shows that increasing number of slices, can also improve the efficiency of resource allocation.

Fig. 6.6 shows the average satisfaction for network slices in different types of network chunks, when there are 30 network slices. The average is calculated as follows,

$$f_{k,m} = \frac{1}{k} \sum_{k=1}^K \frac{\sigma_{k,m}^* - r_{k,m}}{R_{k,m} - r_{k,m}} \quad (6.21)$$

From this figure, it can be seen that the TC-NS provides up to 50% higher slice satisfaction compared to TA-NS and PB-NS. The observation here confirms better performance of the strategy in TC-NS that combines TA-NS revenue maximisation strategy with optimal resource allocation. In PB-NS,

6.6. Performance Investigation

Table 6.1: Jain's Fairness Index of Each Network Chunk

Fairness Index	PB-NS	TA-NS	TC-NS
$C - \xi_1$	0.5858	0.7172	0.8190
$C - \xi_2$	0.5068	0.6193	0.7683
$C - \xi_3$	0.4566	0.6189	0.7444
$C - \xi_4$	0.4092	0.7201	0.8047
$C - \xi_5$	0.3931	0.6109	0.7016
$T - \xi$	0.4911	0.6785	0.8543

network resources have been allocated to slices based on their own requirements without considering global optimisation of resource allocations. Moreover, PB-NS does not consider effective resource allocation, hence, limited network resources will not satisfy the requirements of a portion of slices and this causes decrease in the overall satisfaction.

Table 6.1 lists the Jain's fairness index of allocated resources from each network chunk in different simulation scenarios based on the fairness, expressed in Equation (6.17). From this table, it can be noticed that the Jain's fairness index is the highest, for each chunk, in strategy TC-NS. It means that the each network chunk has fairly allocated its network resources according to resource requirements from each network slices. It is also worth noticing that the fairness of PB-NS decreases from 0.58 (radio chunk) down to 0.39 (central cloud chunk), while TA-NS and TC-NS decrease from 0.71 down to 0.61 and from 0.81 down to 0.7, respectively; there results underline how TA-NS and TC-NS can effectively provide better distribution of fairness among the available network chunks. The Jain's fairness index of the total allocated resource in different simulation strategies based on the fairness Equation (6.18), is also listed in the last row of Table 6.1. It can be seen that fairness index has the highest value in the third scenario, i.e. TC-NS. This underlines that TC-NS is able to increase the network revenue

6.6. Performance Investigation

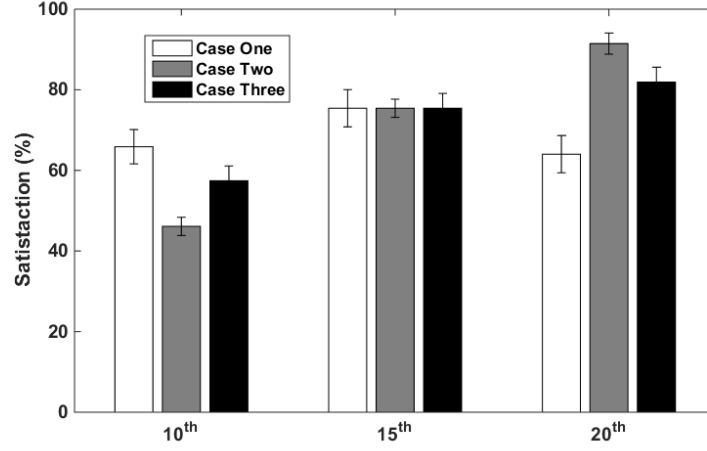


Figure 6.7: Different Requests and Different Priority for network slices for the strategy TC-NS. In case two, the 10th network slice requests 30% more network resources, while the 20th network slice requests 30% less network resources. In case three, the 10th network slice holds 30% higher priority value in case three, while the 20th network slice holds 30% less priority value. The request/priority of the 15th network slice in three cases are the same.)

while considering fairness in the allocation of network resources.

Network Slices with different Requests: In the next step, it focuses the attention on the TC-NS strategy and it is assumed that network slices can have different requests in terms of resources and priority. Three different cases are studied here:

1. All network slices have the same priority value and request the same amount of resources from the network chunks as in the previous simulation results;
2. All network slices have the same priority but have different resource requirements. In this case, 30% more network resources have been requested from slices 1st – 14th, while 30% less network resources have been requested from slices 16th – 30th. The slice 15th has been considered to have the same amount of network resource requirements as in case one.
3. All network slices have the same amount of resource requirements but

6.6. Performance Investigation

with different priority. In this case, higher priority ($\alpha_k = 7$) is considered for slices $1^{st} - 14^{th}$, while lower priority ($\alpha_k = 3$) is considered for slices $16^{th} - 30^{th}$. The slice 15^{th} has the same priority as in case one.

Fig. 6.7 shows satisfaction values for network slices according to their different resource requirements. Observing from Fig. 6.7, the average satisfaction of slice 15^{th} with TC-NS is around 70% (the same in all three cases): this underlines that the performance of our proposed TC-NS strategy in satisfying one slice is not affected by changes in the requests or priorities of other slices. Network slice 10^{th} and 20^{th} are selected as representatives higher and lower resource requests and priorities, respectively. Compared with slice 15^{th} , it can be seen that, if the network slice with higher resource requirements (i.e., slice 10^{th} in case two) receives the same amount of network resource, it will not receive similar average slice satisfaction. Therefore, the network slice with higher resource requirements (i.e., 10^{th} in case two) will increase their payment in order to receive its required network resource and average satisfaction.

Further examination from this simulation shows network slices with higher priority value (e.g., 10^{th} in case three) will receive higher paying price from network chunks, computed based on Equation (6.9) (the revenue values are summarized in Table 6.2). Therefore, the allocated network resources to the service slice k will be less, and it will generate lower satisfaction values, as observed in Fig. 6.7. From this point, the network slice with higher priority value will pay more in order to receive their required network resources and achieve their average satisfaction values.

6.7. Conclusion

Table 6.2: Revenues of Selected Slices of Three Cases in TC-NS

	Case One	Case Two	Case Three
$10^{th} Slice$	0.20	0.35	0.25
$15^{th} Slice$	0.30	0.30	0.30
$20^{th} Slice$	0.27	0.13	0.21

6.7 Conclusion

In this chapter, the economics of network slicing has been studied, how resources can be allocated to the slices competitively and how mobile operators can define revenue model for slicing their network resources for different vertical industries. The proposed approach is an auction-based optimal mechanism that takes into consideration both radio network resources and the storage and computational resource for performing resource allocation. Accordingly, the proposed mechanism can be used to maximise the total network revenue and optimise the resource allocation by considering the limited network virtual resources. Extensive simulations show increase in total network revenue and extra efficiency in allocating network resources to the slices, as well as enhancement in satisfying slices requirements. It also studied further about fairness in allocating resources to the network slices and show how fairness can be improved.

Chapter 7

Conclusions and Future Works

7.1 Conclusions

This thesis examines how these novel mechanisms have been designed and demonstrates how these designed mechanisms affect network performance. In addition, the thesis includes scenarios of demonstrating how network slicing can be implemented in the virtual network environment for enhancing network business model.

Firstly, a fully distributed device centric traffic management mechanism has been proposed by considering their own QoS requirements such as bandwidth demand and battery lifetime requirements. Moreover, the proposed mechanism benefits of the network analytic mechanism by communicating with ANDSF server directly.

Secondly, a novel QFM mechanism, which is composed of a fully distributed mechanism at mobile users side and centralised fairness mechanism at the cloud controller side, has been proposed for improving the efficiency of traffic management and maximising mobile users' quality level values. In the proposed mechanism, how allocating traffic resources in a fairness way is the most significant issue. From the analysis results, users' quality level

7.2. Future Works

has been maximised and network resources have been fairly allocated, the fairness index has been improved.

Thirdly, a novel admission control mechanism of resource allocation in heterogeneous mobile networks has been proposed. The proposed mechanism is heuristic based by considering inter and intra slice priorities. From the analysis results, it increases the quality level of mobile users and allows a better management of network traffic.

Finally, an economics based network slicing mechanism has been proposed in the 5G mobile networks. The complicated mechanism aims to maximise the network revenue and optimise the resource allocation by considering the limited network resources. The proposed mechanism takes into account both radio network resources and the core network resources. From the analysis results, the network revenue and the efficiency of resource allocation have been improved.

The conclusion is that network resources have been efficiently managed but it is still far from complete. This thesis covers some of the key issues in heterogeneous mobile networks and provides foundations for future works. There are still a number of challenges remain that need to be addressed. The following section highlights some research directions for the future.

7.2 Future Works

One direction would be researches of extended network slice allocation problem by considering efficient resource allocation and pricing of network slices. The problem is an NP-hard optimisation problem and could be addressed by investigating a combinatorial auction mechanism in the environment of network virtualization. The combinatorial auction mechanism could focus on improving economic efficiency and enhancing auction revenues of mobile

7.2. Future Works

networks. Following the methods proposed in [162], combinatorial auction linear programming algorithm can be extracted and applied in network virtualisation.

Another direction would be researches of interfacing and control signalling [58, 56]. In wireless network virtualisation, infrastructure providers should provide proper interfaces among themselves and to service providers, so that infrastructure providers can express real-time sharing information and service providers can express their resource requirements for serving end users. Moreover, because delay and jitter effect visualised network greatly and they can be used to evaluate the efficiency of network management, proper control signalling strategies could be designed in order to maintain reliable communication in wireless network virtualisation.

Further researches could also focus on enhancing mobility management from both location profiles and seamless handover aspects in wireless network virtualisation. Efficient mobility management ensures successful delivery of new communications to mobile users and maintains ongoing communication with minimal disruptions. In virtual wireless networks, managing mobile users logically, when they are moving from one resource slice to another in order to satisfy their dynamical changed QoS requirements, could be done further for maintaining service continuity in the future.

In addition, it is worth to investigate a proper security mechanism in wireless network virtualisation. In virtual wireless networks, programmable entities can take advantage of the virtualised mechanism to increase vulnerabilities and threats compared with traditional wireless networks [163]. Therefore, efficient security functions could be studied carefully in the future.

Bibliography

- [1] M. Jo, T. Maksymyuk, R. L. Batista, T. F. Maciel, A. L. F. de Almeida, and M. Klymash, “A survey of converging solutions for heterogeneous mobile networks,” *IEEE Wireless Communications*, vol. 21, no. 6, pp. 54–62, December 2014.
- [2] A. Moubayed, A. Shami, and H. Lutfiyya, “Wireless resource virtualization with device-to-device communication underlaying lte network,” *Broadcasting, IEEE Transactions on*, vol. PP, no. 99, 2015.
- [3] M. Agiwal, A. Roy, and N. Saxena, “Next generation 5g wireless networks: A comprehensive survey,” *IEEE Communications Surveys Tutorials*, vol. PP, no. 99, pp. 1–1, 2016.
- [4] Y. Zhang, C. Lee, D. Niyato, and P. Wang, “Auction approaches for resource allocation in wireless systems: A survey,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 3, Third 2013.
- [5] P. Wang, W. Song, D. Niyato, and Y. Xiao, “Qos-aware cell association in 5g heterogeneous networks with massive mimo,” *IEEE Network*, vol. 29, no. 6, pp. 76–82, Nov 2015.
- [6] E. Hossain and M. Hasan, “5g cellular: key enabling technologies and research challenges,” *IEEE Instrumentation Measurement Magazine*, vol. 18, no. 3, pp. 11–21, June 2015.

BIBLIOGRAPHY

- [7] F. Boccardi, R. W. Heath, A. Lozano, T. L. Marzetta, and P. Popovski, “Five disruptive technology directions for 5g,” *IEEE Communications Magazine*, vol. 52, no. 2, February 2014.
- [8] C. T. Chen, W. C. Su, and C. L. Chen, “Video sharing with seamless service handoff in mobile device-centric cloud computing environment,” in *Consumer Electronics - Taiwan (ICCE-TW), 2015 IEEE International Conference on*, June 2015, pp. 362–363.
- [9] R. Sofia, A. Bogliolo, F. Sivrikaya, and H. Zhu, “User centric networking and services,” *IEEE Communication Magazine*, vol. 52, no. 9, p. 18, 2014.
- [10] W. Lee, M. Bae, H. Kim, S. K. Lee, and H. Kim, “Macod: Mobile device-centric approach for identifying access point congestion,” in *2015 IEEE 29th International Conference on Advanced Information Networking and Applications*, March 2015, pp. 852–859.
- [11] T. Casey and G. M. Muntean, “An integrated data offloading approach for mobile users in urban environments,” in *Signals and Systems Conference (ISSC), 2015 26th Irish*, June 2015, pp. 1–6.
- [12] M. Macuha, D. Amgalan, and T. Derham, “Device-centric approach to improve resiliency of emergency communication,” in *Humanitarian Technology Conference (R10-HTC), 2013 IEEE Region 10*, Aug 2013, pp. 124–129.
- [13] A. Aissioui, A. Ksentini, A. M. Gueroui, and T. Taleb, “Toward elastic distributed sdn/nfv controller for 5g mobile cloud management systems,” *IEEE Access*, vol. 3, pp. 2055–2064, 2015.
- [14] D. Kreutz, F. M. V. Ramos, P. E. Verssimo, C. E. Rothenberg, S. Azodolmolky, and S. Uhlig, “Software-defined networking: A com-

BIBLIOGRAPHY

- prehensive survey,” *Proceedings of the IEEE*, vol. 103, no. 1, pp. 14–76, 2015.
- [15] M. Banikazemi, D. Olshefski, A. Shaikh, J. Tracey, and G. Wang, “Meridian: an sdn platform for cloud network services,” *IEEE Communications Magazine*, vol. 51, no. 2, pp. 120–127, 2013.
- [16] M. Usman, A. A. Gebremariam, U. Raza, and F. Granelli, “A software-defined device-to-device communication architecture for public safety applications in 5g networks,” *IEEE Access*, vol. 3, pp. 1649–1654, 2015.
- [17] A. M. Akhtar, X. Wang, and L. Hanzo, “Synergistic spectrum sharing in 5g hetnets: A harmonized sdn-enabled approach,” *IEEE Communications Magazine*, vol. 54, no. 1, pp. 40–47, 2016.
- [18] S. Lange, S. Gebert, T. Zinner, P. Tran-Gia, D. Hock, M. Jarschel, and M. Hoffmann, “Heuristic approaches to the controller placement problem in large scale sdn networks,” *IEEE Transactions on Network and Service Management*, vol. 12, no. 1, pp. 4–17, 2015.
- [19] F. Giust, L. Cominardi, and C. J. Bernardos, “Distributed mobility management for future 5g networks: overview and analysis of existing approaches,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 142–149, January 2015.
- [20] A. Mendiola, J. Astorga, E. Jacob, M. Higuero, A. Urtasun, and V. Fuentes, “Dynpac: A path computation framework for sdn,” in *2015 Fourth European Workshop on Software Defined Networks*, 2015, pp. 119–120.
- [21] A. Blenk, A. Basta, M. Reisslein, and W. Kellerer, “Survey on network virtualization hypervisors for software defined networking,”

BIBLIOGRAPHY

- IEEE Communications Surveys Tutorials*, vol. 18, no. 1, pp. 655–685, Firstquarter 2016.
- [22] Y. Li and M. Chen, “Software-defined network function virtualization: A survey,” *IEEE Access*, vol. 3, pp. 2542–2553, 2015.
- [23] S. Sun, M. Kadoch, L. Gong, and B. Rong, “Integrating network function virtualization with sdr and sdn for 4g/5g networks,” *IEEE Network*, vol. 29, no. 3, pp. 54–59, May 2015.
- [24] S. Sun, L. Gong, B. Rong, and K. Lu, “An intelligent sdn framework for 5g heterogeneous networks,” *IEEE Communications Magazine*, vol. 53, no. 11, pp. 142–147, November 2015.
- [25] L. Y. Tseng and S. C. Chen, “Two-phase genetic local search algorithm for the multimode resource-constrained project scheduling problem,” *IEEE Transactions on Evolutionary Computation*, vol. 13, no. 4, pp. 848–857, Aug 2009.
- [26] B. A. A. Nunes, M. Mendonca, X. N. Nguyen, K. Obraczka, and T. Turetli, “A survey of software-defined networking: Past, present, and future of programmable networks,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 3, pp. 1617–1634, Third 2014.
- [27] Y. Li and M. Chen, “Software-defined network function virtualization: A survey,” *IEEE Access*, vol. 3, pp. 2542–2553, 2015.
- [28] F. Hu, Q. Hao, and K. Bao, “A survey on software-defined network and openflow: From concept to implementation,” *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, pp. 2181–2206, Fourthquarter 2014.

BIBLIOGRAPHY

- [29] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 27–51, Firstquarter 2015.
- [30] H. H. Cho, C. F. Lai, T. K. Shih, and H. C. Chao, "Integration of sdr and sdn for 5g," *IEEE Access*, vol. 2, pp. 1196–1204, 2014.
- [31] X. Xu, H. Zhang, X. Dai, Y. Hou, X. Tao, and P. Zhang, "Sdn based next generation mobile network with service slicing and trials," *China Communications*, vol. 11, no. 2, pp. 65–77, Feb 2014.
- [32] M. Lahby, S. Baghla, and A. Sekkaki, "Survey and comparison of madm methods for network selection access in heterogeneous networks," in *2015 7th International Conference on New Technologies, Mobility and Security (NTMS)*, July 2015, pp. 1–6.
- [33] V. Sasirekha and M. Ilanzkumaran, "Heterogeneous wireless network selection using fahp integrated with topsis and vikor," in *Pattern Recognition, Informatics and Mobile Engineering (PRIME), 2013 International Conference on*, Feb 2013, pp. 399–407.
- [34] A. Mehbodniya, F. Kaleem, K. K. Yen, and F. Adachi, "A novel wireless network access selection scheme for heterogeneous multimedia traffic," in *2013 IEEE 10th Consumer Communications and Networking Conference (CCNC)*, Jan 2013, pp. 485–489.
- [35] Q. T. N. Vuong, Y. G. Doudane, and N. Agoulmine, "On utility models for access network selection in wireless heterogeneous networks," in *NOMS 2008 - 2008 IEEE Network Operations and Management Symposium*, April 2008, pp. 144–151.
- [36] Q. T. Nguyen-Vuong, N. Agoulmine, E. H. Cherkaoui, and L. Toni, "Multicriteria optimization of access selection to improve the quality

BIBLIOGRAPHY

- of experience in heterogeneous wireless access networks,” *IEEE Transactions on Vehicular Technology*, vol. 62, no. 4, May 2013.
- [37] M. Lahby, L. Cherkaoui, and A. Adib, “An enhanced evaluation model for vertical handover algorithm in heterogeneous networks,” *CoRR*, vol. abs/1206.1848, 2012.
- [38] L. S. mei, P. Su, and X. M. hai, “An improved topsis vertical hand-off algorithm for heterogeneous wireless networks,” in *Communication Technology (ICCT), 2010 12th IEEE International Conference on*, Nov 2010, pp. 750–754.
- [39] F. Bari and V. Leung, “Multi-attribute network selection by iterative topsis for heterogeneous wireless access,” in *2007 4th IEEE Consumer Communications and Networking Conference*, Jan 2007, pp. 808–812.
- [40] S. F. Yang, J.-S. Wu, and H. H. Huang, “A vertical media-independent handover decision algorithm across wi-fi and wimax networks,” in *2008 5th IFIP International Conference on Wireless and Optical Communications Networks (WOCN '08)*, May 2008, pp. 1–5.
- [41] N. Omheni, F. Zarai, M. S. Obaidat, and K. F. Hsiao, “A novel vertical handoff decision making algorithm across heterogeneous wireless networks,” in *Computer, Information and Telecommunication Systems (CITS), 2014 International Conference on*, July 2014, pp. 1–6.
- [42] W. Zhang, “Handover decision using fuzzy madm in heterogeneous networks,” in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, vol. 2, March 2004, pp. 653–658 Vol.2.
- [43] M. S. Dang, A. Prakash, D. K. Anvekar, D. Kapoor, and R. Shorey, “Fuzzy logic based handoff in wireless networks,” in *Vehicular Tech-*

BIBLIOGRAPHY

- nology Conference Proceedings, 2000. VTC 2000-Spring Tokyo. 2000 IEEE 51st*, vol. 3, 2000, pp. 2375–2379 vol.3.
- [44] M. Lahby, L. Cherkaoui, and A. Adib, “An enhanced-topsis based network selection technique for next generation wireless networks,” in *Telecommunications (ICT), 2013 20th International Conference on*, May 2013, pp. 1–5.
- [45] Q. Song and A. Jamalipour, “A network selection mechanism for next generation networks,” in *IEEE International Conference on Communications, 2005. ICC 2005. 2005*, vol. 2, May 2005, pp. 1418–1422 Vol. 2.
- [46] L. Mohamed, C. Leghris, and A. Abdellah, “A hybrid approach for network selection in heterogeneous multi-access environments,” in *New Technologies, Mobility and Security (NTMS), 2011 4th IFIP International Conference on*, Feb 2011, pp. 1–5.
- [47] M. Fathi, V. Maihami, and P. Moradi, “Reinforcement learning for multiple access control in wireless sensor networks: Review, model, and open issues,” *Wirel. Pers. Commun.*, vol. 72, no. 1, pp. 535–547, Sep. 2013.
- [48] H. A. Al-Rawi, M. A. Ng, and K.-L. A. Yau, “Application of reinforcement learning to routing in distributed wireless networks: A review,” *Artif. Intell. Rev.*, vol. 43, no. 3, pp. 381–416, Mar. 2015.
- [49] K.-L. A. Yau, P. Komisarczuk, and P. D. Teal, “Reinforcement learning for context awareness and intelligence in wireless networks: Review, new features and open issues,” *Journal of Network and Computer Applications*, vol. 35, no. 1, pp. 253 – 267, 2012, collaborative Computing and Applications.

BIBLIOGRAPHY

- [50] C. X. Wang, F. Haider, X. Gao, X. H. You, Y. Yang, D. Yuan, H. M. Aggoune, H. Haas, S. Fletcher, and E. Hepsaydir, “Cellular architecture and key technologies for 5g wireless communication networks,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 122–130, February 2014.
- [51] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, “Design considerations for a 5g network architecture,” *IEEE Communications Magazine*, vol. 52, no. 11, pp. 65–75, Nov 2014.
- [52] M. Peng, Y. Li, Z. Zhao, and C. Wang, “System architecture and key technologies for 5g heterogeneous cloud radio access networks,” *IEEE Network*, vol. 29, no. 2, pp. 6–14, March 2015.
- [53] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, “Scaling up mimo: Opportunities and challenges with very large arrays,” *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, Jan 2013.
- [54] A. Gupta and R. K. Jha, “A survey of 5g network: Architecture and emerging technologies,” *IEEE Access*, vol. 3, pp. 1206–1232, 2015.
- [55] S. Chen and J. Zhao, “The requirements, challenges, and technologies for 5g of terrestrial mobile telecommunication,” *IEEE Communications Magazine*, vol. 52, no. 5, pp. 36–43, May 2014.
- [56] N. M. M. K. Chowdhury and R. Boutaba, “Network virtualization: state of the art and research challenges,” *IEEE Communications Magazine*, vol. 47, no. 7, pp. 20–26, July 2009.
- [57] A. Fischer, J. F. Botero, M. T. Beck, H. de Meer, and X. Hesselbach, “Virtual network embedding: A survey,” *IEEE Communications Surveys Tutorials*, vol. 15, no. 4, pp. 1888–1906, Fourth 2013.

BIBLIOGRAPHY

- [58] C. Liang and F. R. Yu, “Wireless network virtualization: A survey, some research issues and challenges,” *IEEE Communications Surveys Tutorials*, vol. 17, no. 1, pp. 358–380, Firstquarter 2015.
- [59] A. Belbekkouche, M. M. Hasan, and A. Karmouch, “Resource discovery and allocation in network virtualization,” *IEEE Communications Surveys Tutorials*, vol. 14, no. 4, pp. 1114–1128, Fourth 2012.
- [60] N. M. K. Chowdhury and R. Boutaba, “A survey of network virtualization,” *Computer Networks*, vol. 54, no. 5, pp. 862 – 876, 2010.
- [61] R. Kokku, R. Mahindra, H. Zhang, and S. Rangarajan, “Nvs: A substrate for virtualizing wireless resources in cellular networks,” *IEEE/ACM Transactions on Networking*, vol. 20, no. 5, pp. 1333–1346, Oct 2012.
- [62] Y. Li, W. Li, and C. Jiang, “A survey of virtual machine system: Current technology and future trends,” in *Electronic Commerce and Security (ISECS), 2010 Third International Symposium on*, July 2010, pp. 332–336.
- [63] Z. Zhang, Z. Li, K. Wu, D. Li, H. Li, Y. Peng, and X. Lu, “Vmthunder: Fast provisioning of large-scale virtual machine clusters,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3328–3338, Dec 2014.
- [64] I. da Silva, G. Mildh, A. Kaloxylos, P. Spapis, E. Buracchini, A. Trogolo, G. Zimmermann, and N. Bayer, “Impact of network slicing on 5g radio access networks,” in *2016 European Conference on Networks and Communications (EuCNC)*, June 2016, pp. 153–157.
- [65] M. R. Sama, X. An, Q. Wei, and S. Beker, “Reshaping the mobile core network via function decomposition and network slicing for the 5g era,”

BIBLIOGRAPHY

- in *2016 IEEE Wireless Communications and Networking Conference*, 2016, pp. 1–7.
- [66] K. Katsalis, N. Nikaein, E. Schiller, R. Favraud, and T. I. Braun, “5g architectural design patterns,” in *2016 IEEE International Conference on Communications Workshops (ICC)*, 2016, pp. 32–37.
- [67] M. Yu, J. Rexford, X. Sun, S. Rao, and N. Feamster, “A survey of virtual lan usage in campus networks,” *IEEE Communications Magazine*, vol. 49, no. 7, pp. 98–103, July 2011.
- [68] X. Xiao, A. Hannan, B. Bailey, and L. M. Ni, “Traffic engineering with mpls in the internet,” *IEEE Network*, vol. 14, no. 2, pp. 28–33, Mar 2000.
- [69] S. B. Rejeb, N. Nasser, and S. Tabbane, “Admission control strategies and qos evaluation based on mobility in lte,” in *2013 IEEE Global Communications Conference (GLOBECOM)*, Dec 2013, pp. 4883–4888.
- [70] S. Parsaeefard, V. Jumba, M. Derakhshani, and T. Le-Ngoc, “Joint resource provisioning and admission control in wireless virtualized networks,” in *Wireless Communications and Networking Conference (WCNC), 2015 IEEE*, March 2015.
- [71] A. Ksentini, T. Taleb, and K. B. Letaif, “Qoe-based flow admission control in small cell networks,” *IEEE Transactions on Wireless Communications*, vol. 15, no. 4, pp. 2474–2483, April 2016.
- [72] M. Z. Chowdhury and Y. M. Jang, “Handover management in high-dense femtocellular networks,” *EURASIP Journal on Wireless Communications and Networking*, vol. 6, January 2013.

BIBLIOGRAPHY

- [73] L. B. Le, D. T. Hoang, D. Niyato, E. Hossain, and D. I. Kim, "Joint load balancing and admission control in ofdma-based femtocell networks," in *2012 IEEE International Conference on Communications (ICC)*, June 2012, pp. 5135–5139.
- [74] L. B. Le, D. Niyato, E. Hossain, D. I. Kim, and D. T. Hoang, "Qos-aware and energy-efficient resource management in ofdma femtocells," *IEEE Transactions on Wireless Communications*, vol. 12, no. 1, pp. 180–194, January 2013.
- [75] M. Monemi, A. R. Z. Asli, S. G. Haghighi, and M. Rasti, "Distributed multiple target-sinrs tracking power control in wireless multirate data networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1850–1859, April 2013.
- [76] Y. F. Liu, Y. H. Dai, and S. Ma, "Joint power and admission control: Non-convex l approximation and an effective polynomial time deflation approach," *IEEE Transactions on Signal Processing*, vol. 63, no. 14, July 2015.
- [77] M. Monemi, M. Rasti, and E. Hossain, "Low-complexity sinr feasibility checking and joint power and admission control in prioritized multitier cellular networks," *IEEE Transactions on Wireless Communications*, vol. 15, no. 3, pp. 2421–2434, March 2016.
- [78] E. Monemi, Mehdi Rasti and Hossain, "On low-complexity sinr feasibility checking and joint power and admission control in prioritized multi-tier cellular networks under co-channel deployment," *Wireless Communications, IEEE Transactions on*, vol. PP, no. 99, 2015.
- [79] X. Zhai, L. Zheng, and C. W. Tan, "Energy-infeasibility tradeoff in cognitive radio networks: Price-driven spectrum access algorithms,"

BIBLIOGRAPHY

- IEEE Journal on Selected Areas in Communications*, vol. 32, no. 3, March 2014.
- [80] C. A. Gizelis and D. D. Vergados, “A survey of pricing schemes in wireless networks,” *IEEE Communications Surveys Tutorials*, vol. 13, no. 1, First 2011.
- [81] N. Khan, M. G. Martini, and Z. Bharucha, “Quality-aware fair down-link scheduling for scalable video transmission over lte systems,” in *2012 IEEE 13th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2012, pp. 334–338.
- [82] B. Wang, Z. Han, and K. J. R. Liu, “Distributed relay selection and power control for multiuser cooperative communication networks using buyer/seller game,” in *IEEE INFOCOM 2007 - 26th IEEE International Conference on Computer Communications*, May 2007.
- [83] N. Nie and C. Comaniciu, “Adaptive channel allocation spectrum etiquette for cognitive radio networks,” in *First IEEE International Symposium on New Frontiers in Dynamic Spectrum Access Networks, 2005. DySPAN 2005.*, Nov 2005.
- [84] X. Kang, Y. C. Liang, and H. K. Garg, “Distributed power control for spectrum-sharing femtocell networks using stackelberg game,” in *2011 IEEE International Conference on Communications (ICC)*, June 2011.
- [85] P. Semasinghe, E. Hossain, and K. Zhu, “An evolutionary game for distributed resource allocation in self-organizing small cells,” *IEEE Transactions on Mobile Computing*, vol. 14, no. 2, Feb 2015.
- [86] M. Bennis, S. Guruacharya, and D. Niyato, “Distributed learning strategies for interference mitigation in femtocell networks,” in *Global*

BIBLIOGRAPHY

- Telecommunications Conference (GLOBECOM 2011)*, 2011 IEEE, Dec 2011.
- [87] A. Maaref, J. Ma, M. Salem, H. Baligh, and K. Zarifi, “Device-centric radio access virtualization for 5g networks,” in *Globecom Workshops (GC Wkshps)*, 2014, Dec 2014.
- [88] S. Andreev, M. Gerasimenko, O. Galinina, Y. Koucheryavy, N. Himayat, S.-P. Yeh, and S. Talwar, “Intelligent access network selection in converged multi-radio heterogeneous networks,” *Wireless Communications, IEEE*, vol. 21, no. 6, December 2014.
- [89] “3gpp ts 24.312 access network discovery and selection function (andsf) management object (mo),” *Release 12*, vol. v12.8.0, 2015.
- [90] “Access network discovery and selection function (ADNSF) management object (MO),” 3GPP TS 24.312, September 2014.
- [91] Y. M. Kwon, J. S. Kim, J. Gu, and M. Y. Chung, “Andsf-based congestion control procedure in heterogeneous networks,” in *The International Conference on Information Networking 2013 (ICOIN)*, 2013, pp. 547–550.
- [92] A. Kaloxylos, I. Modeas, F. Georgiadis, and N. Passas, “Network Selection Algorithm for Heterogeneous Wireless Networks: from Design to Implementation,” *Network Protocols and Algorithms*, vol. 1, 2009.
- [93] M. Lahby, L. Cherkaoui, and A. Adib, “A Novel Ranking Algorithm Based Network Selection For Heterogeneous Wireless Access,” *Journal of Networks*, vol. 8, Feb. 2013.
- [94] “Multi-RAT Traffic Steering,” Nomor Research White Paper, December 2011.

BIBLIOGRAPHY

- [95] Q.-T. Nguyen-Vuong, N. Agoulmine, E. Cherkaoui, and L. Toni, "Multicriteria optimization of access selection to improve the quality of experience in heterogeneous wireless access networks," *Vehicular Technology, IEEE Transactions on*, vol. 62, no. 4, May 2013.
- [96] M. Louta and P. Bellavista, "Bringing always best connectivity vision a step closer: challenges and perspectives," *IEEE Communications Magazine*, vol. 51, no. 2, February 2013.
- [97] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, June 2011.
- [98] "Access Network Discovery and Selection Function (ANDSF) Management Object (MO)," 3GPP TS 24.312, September 2014.
- [99] D. Lopez-Perez, I. Guvenc, G. de la Roche, M. Kountouris, T. Quek, and J. Zhang, "Enhanced intercell interference coordination challenges in heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, June 2011.
- [100] L. Li, X. Hu, and W. Zhang, "Design of an arm-based power meter having wifi wireless communication module," in *Industrial Electronics and Applications, 2009. ICIEA 2009. 4th IEEE Conference on*, May 2009.
- [101] J. Li, J. Xiao, J. W. K. Hong, and R. Boutaba, "Application-centric wi-fi energy management on smart phone," in *Network Operations and Management Symposium (APNOMS), 2012 14th Asia-Pacific*, Sept 2012.

BIBLIOGRAPHY

- [102] C. Ma and Y. Yang, “A battery-aware scheme for routing in wireless ad hoc networks,” *Vehicular Technology, IEEE Transactions on*, vol. 60, no. 8, Oct 2011.
- [103] J. Cho, Y. Woo, S. Kim, and E. Seo, “A battery lifetime guarantee scheme for selective applications in smart mobile devices,” *Consumer Electronics, IEEE Transactions on*, vol. 60, no. 1, February 2014.
- [104] R. Trestian, O. Ormond, and G.-M. Muntean, “Energy quality cost tradeoff in a multimedia-based heterogeneous wireless network environment,” *Broadcasting, IEEE Transactions on*, vol. 59, no. 2, June 2013.
- [105] I. Bisio, M. Cerruti, F. Lavagetto, M. Marchese, M. Pastorino, A. Randazzo, and A. Sciarrone, “A trainingless wifi fingerprint positioning approach over mobile devices,” *IEEE Antennas and Wireless Propagation Letters*, vol. 13, 2014.
- [106] Q.-T. Nguyen-Vuong, Y. Ghamri-Doudane, and N. Agoulmine, “On utility models for access network selection in wireless heterogeneous networks,” in *Network Operations and Management Symposium, 2008. NOMS 2008. IEEE*, April 2008.
- [107] S. M. Alavi, C. Zhou, and W. W. Gen, “Efficient and fair resource allocation scheme for ofdma networks based on auction game,” in *Vehicular Technology Conference (VTC Fall), 2012 IEEE*, Sept 2012.
- [108] P. Muoz, D. Laselva, R. Barco, and P. Mogensen, “Dynamic traffic steering based on fuzzy q-learning approach in a multi-rat multi-layer wireless network,” *Computer Networks*, vol. 71, no. 0, pp. 100 – 116, 2014.

BIBLIOGRAPHY

- [109] A. Saleh, S. Redana, B. Raaf, and J. Hamalainen, "Comparison of relay and pico enb deployments in lte-advanced," in *Vehicular Technology Conference Fall (VTC 2009-Fall)*, 2009 IEEE 70th, Sept 2009.
- [110] Y. Bai, J. Zhou, and L. Chen, "Hybrid spectrum usage for overlaying lte macrocell and femtocell," in *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, Nov 2009.
- [111] Y. Zaki, T. Weerawardane, S. Hauth, E. Wallmeier, and C. Grg, "Intelligent traffic enforcement for lte backhaul," in *2013 IEEE 24th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC)*, Sept 2013.
- [112] H. Elshaer, F. Boccardi, M. Dohler, and R. Irmer, "Downlink and uplink decoupling: a disruptive architectural design for 5g networks," *CoRR*, vol. abs/1405.1853, 2014.
- [113] R. Guerzoni, R. Trivisonno, and D. Soldani, "Sdn-based architecture and procedures for 5g networks," in *5G for Ubiquitous Connectivity (5GU), 2014 1st International Conference on*, Nov 2014.
- [114] T. Mahmoodi and S. Seetharaman, "Traffic jam: Handling the increasing volume of mobile data traffic," *IEEE Vehicular Technology Magazine*, vol. 9, no. 3, Sept 2014.
- [115] P. Ameigeiras, J. J. Ramos-munoz, L. Schumacher, J. Prados-Garzon, J. Navarro-Ortiz, and J. M. Lopez-soler, "Link-level access cloud architecture design based on sdn for 5g networks," *IEEE Network*, vol. 29, no. 2, March 2015.
- [116] "Cisco visual networking index:global mobile data traffic forecast update 20142019," *Cisco White paper*, 2015.

BIBLIOGRAPHY

- [117] D. Tuncer, M. Charalambides, S. Clayman, and G. Pavlou, “Adaptive resource management and control in software defined networks,” *IEEE Transactions on Network and Service Management*, vol. 12, no. 1, March 2015.
- [118] R. Jain, D. Chiu, and W. Hawe, “A quantitative measure of fairness and discrimination for resource allocation in shared computer systems,” *CoRR*, vol. cs.NI/9809099, 1998.
- [119] O. H. Toktam Mahmoodi, V. Friderikos and H. Aghvami, “Cross-layer optimization to maximize fairness among tcp flows of different tcp flavors,” in *IEEE GLOBECOM 2008 - 2008 IEEE Global Telecommunications Conference*, Nov 2008.
- [120] T. Mahmoodi, V. Friderikos, O. Holland, and H. Aghvami, “Tcp-aware resource allocation in ofdma based wireless networks,” in *Cross Layer Design, 2009. IWCLD '09. Second International Workshop on*, June 2009.
- [121] D. Cheng, J. Rao, C. Jiang, and X. Zhou, “Elastic power-aware resource provisioning of heterogeneous workloads in self-sustainable datacenters,” *IEEE Transactions on Computers*, vol. 65, no. 2, Feb 2016.
- [122] C. Guo, M. Sheng, X. Wang, and Y. Zhang, “Throughput maximization with short-term and long-term jain’s index constraints in downlink ofdma systems,” *IEEE Transactions on Communications*, vol. 62, no. 5, May 2014.
- [123] P. Munoz, D. Laselva, R. Barco, and P. Mogensen, “Dynamic traffic steering based on fuzzy q-learning approach in a multi-rat multi-layer wireless network,” *Comput. Netw.*, vol. 71, Oct. 2014.

BIBLIOGRAPHY

- [124] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, “What will 5g be?” *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, June 2014.
- [125] H.-J. Einsiedler, A. Gavras, P. Sellstedt, R. Aguiar, R. Trivisonno, and D. Lavaux, “System design for 5G converged networks,” in *Networks and Communications (EuCNC), 2015 European Conference on*, 2015.
- [126] Alliance, N. G. M. N., “5G white paper. Next Generation Mobile Networks,” White Paper, 2015.
- [127] Y. Zaki, L. Zhao, C. Goerg, and A. Timm-Giel, “Lte wireless virtualization and spectrum management,” in *Wireless and Mobile Networking Conference (WMNC), 2010 Third Joint IFIP*, Oct 2010.
- [128] C. Liang, F. R. Yu, and X. Zhang, “Information-centric network function virtualization over 5g mobile wireless networks,” *IEEE Network*, vol. 29, no. 3, May 2015.
- [129] K. Zhu and E. Hossain, “Virtualization of 5g cellular networks as a hierarchical combinatorial auction,” *CoRR*, vol. abs/1511.08256, 2015.
- [130] X. Costa-Perez, J. Swetina, T. Guo, R. Mahindra, and S. Rangarajan, “Radio access network virtualization for future mobile carrier networks,” *IEEE Communications Magazine*, vol. 51, no. 7, July 2013.
- [131] V. Jumba, S. Parsaeefard, M. Derakhshani, and T. Le-Ngoc, “Energy-efficient robust resource provisioning in virtualized wireless networks,” in *Ubiquitous Wireless Broadband (ICUWB), 2015 IEEE International Conference on*, Oct 2015.

BIBLIOGRAPHY

- [132] Y. L. Lee, T. C. Chuah, J. Loo, and A. Vinel, "Recent advances in radio resource management for heterogeneous lte/lte-a networks," *IEEE Communications Surveys Tutorials*, vol. 16, no. 4, Fourthquarter 2014.
- [133] M. Kamel, L. B. Le, and A. Girard, "Lte wireless network virtualization: Dynamic slicing via flexible scheduling," in *Vehicular Technology Conference (VTC Fall), 2014 IEEE 80th*, Sept 2014.
- [134] M. Kalil, A. Shami, and Y. Ye, "Wireless resources virtualization in lte systems," in *Computer Communications Workshops (INFOCOM WKSHPS), 2014 IEEE Conference on*, April 2014.
- [135] S. Ben Rejeb, N. Nasser, and S. Tabbane, "Admission control strategies and qos evaluation based on mobility in lte," in *Global Communications Conference (GLOBECOM), 2013 IEEE*, Dec 2013.
- [136] "View on 5G Architecture," 5GPPP Architecture Working Group, White Paper, 5G PPP, July 2016.
- [137] H. J. Einsiedler, A. Gavras, P. Sellstedt, R. Aguiar, R. Trivisonno, and D. Lavaux, "System design for 5g converged networks," in *Networks and Communications (EuCNC), 2015 European Conference on*, June 2015.
- [138] Y. Li, L. Liu, H. Li, J. Zhang, and Y. Yi, "Resource allocation for delay-sensitive traffic over lte-advanced relay networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 8, Aug 2015.
- [139] P. K. Agyapong, M. Iwamura, D. Staehle, W. Kiess, and A. Benjebbour, "Design considerations for a 5g network architecture," *IEEE Communications Magazine*, vol. 52, no. 11, Nov 2014.

BIBLIOGRAPHY

- [140] D. Niyato and E. Hossain, “Competitive pricing for spectrum sharing in cognitive radio networks: Dynamic game, inefficiency of nash equilibrium, and collusion,” *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 1, Jan 2008.
- [141] S. Y. Jung, S. M. Yu, and S. L. KIM, “Optimization of spectrum allocation and subsidization in mobile communication services,” *IEEE Transactions on Vehicular Technology*, vol. PP, no. 99, 2015.
- [142] J. Martyna, “Oligopoly bertrand model for price competition in cognitive radio networks,” in *Communication Systems, Networks Digital Signal Processing (CSNDSP), 2014 9th International Symposium on*, July 2014.
- [143] P. Lin, X. Feng, and Q. Zhang, “Flexauc: Serving dynamic demands in spectrum trading markets with flexible auction,” in *INFOCOM, 2014 Proceedings IEEE*, April 2014.
- [144] J. F. Gomes, P. Ahokangas, and S. Moqaddamerad, “Business modeling options for distributed network functions virtualization: Operator perspective,” in *European Wireless 2016; 22th European Wireless Conference*, May 2016.
- [145] X. Zhou, R. Li, T. Chen, and H. Zhang, “Network slicing as a service: enabling enterprises’ own software-defined cellular networks,” *IEEE Communications Magazine*, vol. 54, no. 7, July 2016.
- [146] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, “Network function virtualization in 5g,” *IEEE Communications Magazine*, vol. 54, no. 4, April 2016.

BIBLIOGRAPHY

- [147] M. C. Menglan Jiang and T. Mahmoodi, “Network slicing management and prioritization in 5g mobile systems,” in *European Wireless, 2016*, May 2016.
- [148] H. Zhang, C. Jiang, N. C. Beaulieu, X. Chu, X. Wang, and T. Q. S. Quek, “Resource allocation for cognitive small cell networks: A cooperative bargaining game theoretic approach,” *IEEE Transactions on Wireless Communications*, vol. 14, no. 6, June 2015.
- [149] D. Niyato and E. Hossain, “Competitive spectrum sharing in cognitive radio networks: a dynamic game approach,” *IEEE Transactions on Wireless Communications*, vol. 7, no. 7, pp. 2651–2660, July 2008.
- [150] X. Kang, R. Zhang, and M. Motani, “Price-based resource allocation for spectrum-sharing femtocell networks: A stackelberg game approach,” in *Global Telecommunications Conference (GLOBECOM 2011)*, 2011 IEEE, Dec 2011, pp. 1–5.
- [151] S. Guruacharya, D. Niyato, D. I. Kim, and E. Hossain, “Hierarchical competition for downlink power allocation in ofdma femtocell networks,” *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, April 2013.
- [152] V. N. Ha and L. B. Le, “Distributed base station association and power control for heterogeneous cellular networks,” *IEEE Transactions on Vehicular Technology*, vol. 63, no. 1, Jan 2014.
- [153] J. Jia, Q. Zhang, Q. Zhang, and M. Liu, “Revenue generation for truthful spectrum auction in dynamic spectrum access,” in *Proceedings of the Tenth ACM International Symposium on Mobile Ad Hoc Networking and Computing*, ser. MobiHoc ’09. New York, NY, USA: ACM, 2009.

BIBLIOGRAPHY

- [154] Z. Suli, M. Zhongjing, and L. Xiangdong, “Efficient distributed resource allocation under synchronous auction-based algorithm,” in *Control Conference (CCC), 2015 34th Chinese*, July 2015.
- [155] F. Fu and U. C. Kozat, “Stochastic game for wireless network virtualization,” *IEEE/ACM Transactions on Networking*, vol. 21, no. 1, Feb 2013.
- [156] P. Samadi, H. Mohsenian-Rad, R. Schober, and V. W. S. Wong, “Advanced demand side management for the future smart grid using mechanism design,” *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1170–1180, Sept 2012.
- [157] H. Li, C. Wu, Z. Li, and F. C. M. Lau, “Virtual machine trading in a federation of clouds: Individual profit and social welfare maximization,” *IEEE/ACM Transactions on Networking*, vol. 24, no. 3, pp. 1827–1840, June 2016.
- [158] L. Mashayekhy, M. M. Nejad, D. Grosu, and A. V. Vasilakos, “An online mechanism for resource allocation and pricing in clouds,” *IEEE Transactions on Computers*, vol. 65, no. 4, pp. 1172–1184, April 2016.
- [159] X. Wang, X. Wang, H. Che, K. Li, M. Huang, and C. Gao, “An intelligent economic approach for dynamic resource allocation in cloud services,” *IEEE Transactions on Cloud Computing*, vol. 3, no. 3, pp. 275–289, July 2015.
- [160] T. Mahmoodi, V. Friderikos, O. Holland, and H. Aghvami, “Cross-layer optimization to maximize fairness among tcp flows of different tcp flavors,” in *IEEE Global Telecommunications Conference (GLOBECOM)*, Nov 2008.

BIBLIOGRAPHY

- [161] T. Mahmoodi, V. Friderikos, O. Holland, and A. H. Aghvami, “Optimal design of forward error correction for fairness maximisation among transmission control protocol flavours over wireless networks,” *IET Communications*, vol. 4, no. 10, July 2010.
- [162] S. Zaman and D. Grosu, “Combinatorial auction-based allocation of virtual machine instances in clouds,” in *Cloud Computing Technology and Science (CloudCom), 2010 IEEE Second International Conference on*, Nov 2010, pp. 127–134.
- [163] S. Bu, F. R. Yu, X. P. Liu, and H. Tang, “Structural results for combined continuous user authentication and intrusion detection in high security mobile ad-hoc networks,” *IEEE Transactions on Wireless Communications*, vol. 10, no. 9, pp. 3064–3073, September 2011.